602223

TT-60-23871-1

F- TS-9811/V

TRANSLATION

THE THEORY OF RANDOM PROCESSES AND ITS APPLICATION IN RADIO ENGINEERING

By B. R. Levin

March 1960

475 Pages

602223 (PART I of II, pages 1 - 253)

PRICES SUBJECT TO CHANGE

PREPARED BY
LIAISON OFFICE
TECHNICAL INFORMATION CENTER
MCLTD
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

F- TS-9811/V

Best
Available
Copy

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION # 602223

☒ 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY.

☒ 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT.

☐ 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY.

☐ 4. THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.

☐ 5. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED, DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.

☐ 6. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.

☐ 7. DOCUMENT IS AVAILABLE IN MICROFICHE ONLY.

☐ 8. DOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT DOCUMENTS ONLY).

☐ 9.

NBS 9/64

PROCESSOR: pp

B. R. Levin

TEORIYA SLUCHAYNYKH PROTSESSOV I EE PRIMENENIYE B RADIOTECHNIKE

Izdatel'stvo "Sovetskoye Radio"

Moscow 1957

Foreign Pages: 496

F-TS-9811/V

INSTITUTE FOR APPLIED TECHNOLOGY
EVALUATION REPORT

AD 602223
PB ______
Other ______
Travel No. ______

ROUTING

1) Name G. SHAPIRO 2) Name ______ 3) Name ______
Room 310 Bldg. EL Room ______ Bldg. ______ Room ______ Bldg. ______

INSTRUCTIONS: 1. If incorrectly routed, please strike out your name and forward to a more qualified NBS specialist.
2. If evaluation not possible within 3 days, strike out your name and return to sender, and place in "out" basket.

SENDER: S. ABRAHAMS, ROOM 1220-IND, EXT. 7102

THE FOLLOWING ACTION TO BE TAKEN BY EVALUATOR

Signature of Evaluator J Kawedas Date ______
Div. ______

Check Appropriate Item:

	Yes	No
1. For Publicity		✓
2. Copy for Evaluator		✓

Note: If item 1 above checked "no", sections A,B,C & D need not be completed.

Notice: Comments furnished in this Evaluation Report will be used primarily as a guide for promotional purposes only.

A/ Indicate subject category:

B/ Brief Evaluation or Comments: (Continue on back if necessary)

C/ Indicate Fields and/or Areas of Interest:

D/ Significant Features: (Check if appropriate)
1. State of Art ____
2. Bibliography ____
3. Leading Scientist in Field ____
4. New Material ____

DO NOT WRITE BELOW THIS LINE

	Yes	No
For collection	✓	
For OTR		

IAT Controller ______
Date Received SEP 1 1964
Date Dispatched October 20, 1964

Disposition of Report: To NBS ____ To Other Agency ______ (Specify)

Stock ______

Special Instructions:

Form Pending Approval

USCOMM-NBS-DC

TABLE OF CONTENTS

Contents:

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

FOREWORD

The solution of a large number of problems which are of practical importance in various fields of radio engineering (radio communications, radar, radio navigation, automation and remote control, electronic computers, etc.) is impossible without a more thorough and expanded knowledge of mathematics in the domain of the theory of probabilities in general and in the domain of the theory of random processes in particular.

There exist a number of good systematic courses on probability theory. However, in his practical work an engineer lacks, as a rule, the opportunity to study such a course in its entirety. Furthermore, the presentation of the material in such books, while sufficiently sequential and strict in the mathematical sense, is not aimed at solving concrete engineering problems, and does not contain the technical examples necessary for the radio engineer, which could serve as specimens in the solution of a large number of problems. Up to now, there are no textbooks with a systematic exposition of the mathematical theory of random processes, in a form accessible to radio engineers.

There exists a need for a textbook in which, along with the necessary minimum of information on the theory of random variables and random processes, considerable space would be devoted to carefully selected and worked-out radio-engineering examples which illustrate the possibilities made available to radio engineers, in the investigation of a number of problems, by the use of probability-theory methods.

In attempting to satisfy to a certain extent the requests of radio specialists for such a textbook, the author considers the basic purpose of the present book to be such a presentation of probability-theory methods, applicable to radio engineering, as would assist the engineer in making use of these methods in his professional work. The material presented makes no claim to completeness with respect to mathematical results, and especially not with respect to the physical treatment

of such results. In several places it is necessary to sacrifice mathematical strictness in favor of a more accessible presentation, the solitary purpose of which is to provide engineers with the necessary technical apparatus.

As a result of the limited scope of the book it has also been impossible to devote the necessary attention to the physical treatment of results, which were regarded merely as illustrations of mathematical methods.

The basis of the present book consists of a lecture course on the theory of random processes which has, during the past several years, been delivered by the author to specialist-candidates in radio engineering.

It is assumed that besides the fundamentals of differential and integral calculus, the reader has a command of classical harmonic analysis (series and Fourier integral) within the scope of the usual introductory sections included in many monographs on radio engineering*.

Extensive use is made of special functions, the properties of some of which (Bessel and Gamma functions) are assumed to be familiar to the reader, while the properties of the others (hypergeometric functions and Hermite polynomials) appear in the appendices to this book.

The book has eleven chapters. In the beginning are formulated the rules for the addition and multiplication of probabilities, and examples of practical application of these rules are given, as well as certain of their important consequences. In the second and third chapters are examined distribution functions of one random variable and of groups of random variables, their numerical characteristics, and also formulas for the transformation of distribution functions with the functional transformations of continuous random variables. Particular attention is paid to the method of characteristic functions, which is effectively employed for determining the distribution functions of a sum of random variables.

---

* Cf, for instance, I. S. Gonorovskiy, Radiosignaly i perekhodnyye yavleniya v radiotsepyakh (Radio Signals and Transient Phenomena in Radio Circuits), Moskva, Svyaz'izdat (Communications Publishing House), 1954.

The presentation of the central limiting theorem and of the rule of large numbers is not restricted to only a proof of the limiting relationships. An estimate is given of the rate of convergence of the sums of independent random variables to the normal law. There is introduced the relationship linking the error and reliability of mesaurements in a finite number of experiments.

The fourth chapter completes the round of questions connected with the theory of random variables, which is essential to an understanding of the succeeding six chapters which deal with the theory of random processes and its radio-engineering applications. A number of the examples used in the first four chapters have been selected with a view to the problems examined subsequently.

In the fifth chapter is discussed the basic aspects of the theory of random processes: steady state and ergodicity, the correlation function and its properties, the energy spectrum of a random process and its link with the correlation function, the derivative of a stationary random process, the average number of overshoots of the random process. There is examined the most important form of the random process - the normal random process. Further (Chapter VI) is presented general methods of determining statistical characteristics with the functional transformations of random processes. The following chapters, VII and VIII, illustrate non-linear transformations by the example of the normal random process, and Chapter IX illustrates the transformations of the distribution function of this random process in terms of an electronic circuit: the i-f amplifier, the square-law detector, and the video filter.

In Chapter X is studied the energy spectra of pulse-type random processes, and of continuous oscillations modulated by amplitude and by phase (frequency) by random processes.

The eleventh, and concluding chapter, in which is contained the elements of the theory of information, has been written with the aim of drawing the reader's attention to this furiously developing branch of the theory of random processes and of awakening in him the desire to study it in specialized literature.

The literature on the theory and applications of random processes is very ex-

tensive; it was, therefore, necessary to limit ourselves to citing only a small number of sources, recommended for a more thorough study of individual question. These references are provided at the end of each chapter.

In order not to repeat the definitions of the symbols that have been accepted, they have been listed at the end of the book in front of the index.

The author considers it his pleasant duty to express sincere thanks to Professor I. S. Gonorovskiy and to Candidate E. V. Drakin, whose valuable remarks have been taken into consideration in the preparation of this book for the press.

Chapter I

FUNDAMENTAL ASPECTS OF THE THEORY OF PROBABILITY

1. Definitions and Terminology

For a precise description of the processes taking place in electrical or mechanical systems with one or more degrees of freedom, it is usually sufficient to solve a differential equation, or a system of such equations given the necessary number of initial data (in an electrical system, the voltages, currents, or charges at the instant of time assumed as the beginning of reckoning). Depending on their complexity, these equations are solved by various mathematical methods (for instance, in the case of linear systems, by the operational method).

The difficulty of using the differential-equation apparatus increases sharply in the investigation of physical phenomena with an enormous number of degrees of freedom, for instance in the investigation of the motion of interacting elementary particles of which substances are constituted. These difficulties are not only computational, to a large extent they are linked with the impossibility, in such systems, of assigning exact values to the necessary quantity of initial data (for instance, the position and velocity of the elementary particles of a substance at a fixed moment in time). It becomes necessary to abandon an exact description of the behavior of a system with a large number of degrees of freedom.

However, in the study of systems with a large number of degrees of freedom there become apparent rules of a special type, which are stipulated by just this large number of degrees of freedom. The quantitative increase in the number of degrees of freedom leads to the appearance of qualitatively new rules. These are called statistical rules. The study of statistical rules constitutes the subject of

the theory of probabilities.

The abandonment of exact description signifies the impossibility of predicting the results of individual experiments, which can vary from one experiment to another in an arbitrary manner. Such experiments, for which a precise prediction of observation results in each individual experiment is impossible, are called random ones. An example of a random experiment is the observation of noise in radio circuits. The term "random" refers not to the conditions of the experiment, which must remain unchanged, but to the results of observation in the multifold repetition of the same experiment (i.e., in the realization of a definite complex of conditions). The observation result may be registered as the occurrence or non-occurrence of a certain event A, which is called a random event. Thus, as a result of observing noise in a radio receiver, there may occur or not occur a random event consisting of the fact that the amplitude of the noise will exceed the assigned voltage level U_0.

It is impossible accurately to predict the result of each individual random experiment. But if a succession of a large number of experiments is examined in its entirety, an important statistical rule will appear: the average results in the observation of a sufficiently long series of experiments remain practically constant. This constancy in the average results of a large number of random experiments is of a stable nature.

In n random experiments, let the random occurence A occur m times. The ratio

$$v = \frac{m}{n} \tag{1.1}$$

is called the relative frequency of the occurrence of event A with n random experiments. To predict the occurrence and non-occurrence of A in an individual experiment is impossible. But in observations of the frequency ν of the occurrence of event A, as the number n is increased there is disclosed the property of this frequency to deviate ever less from a certain constant.

Thus, for instance, it is impossible to predict in advance whether a given radio-engineering device (tube, unit, station) will operate without damage for not

less than T hours. However, in the testing of a large number of identical devices, the frequency of a random event, which consists of the fact that the device has operated not less than T hours, will deviate ever less from a certain constant as the number of devices tested is increased.

Thus it is possible to state the following assumption: for each random event A there is a certain constant p to which the frequencies of occurrence ν of the event A prove to be closer the greater the number n of random experiments. The assumption of the existence of such a constant is the experimental basis of probability theory. This constant p is called the probability of the occurrence of the random event A in the realization of a definite complex of conditions.

For the frequency of occurrence of an event there holds true the obvious inequality $0 \leqslant \nu \leqslant 1$. Since, according to the definition, every probability p is approximately equal to a certain frequency ν, it is natural to consider that p satisfies the same inequality

$$0 \leqslant p \leqslant 1. \tag{1.2}$$

An event which occurs invariably in a random experiment is called certain. The probability of a certain event is equal to 1.

An event which cannot occur in a random experiment is called impossible. The probability of an impossible event is equal to zero.

Two events are called incompatible, if they cannot both occur with the realization of a definite complex of conditions.

If it is known for certain that with every random experiment there is observed one of n mutually exclusive events $A_1, A_2, \ldots A_n$, then the indicated aggregate of n events is called a set.

Converse random events are two incompatible events which constitute a set. The simplest example of converse events are certain and impossible events.

If the probability of occurrence of one event A depends on whether or not another event B has occurred, then two such events are called dependent events.

Independent random events are two events for which the probability of occurence of one does not depend on whether or not the other has taken place.

Equiprobable events are called random events which have an equal probability of occurrence.

Two basic rules must be adhered to with respect to probabilities of random events: the addition rule and the multiplication rule, from which a number of important consequences arise. The following paragraphs are devoted to a presentation of these rules and to several examples of their practical application.

2. The Addition Rule.

If $A_1, A_2, \ldots A_n$ are mutually incompatible events, the probability of occurrence of one of the events A_1, or $A_2, \ldots$ or A_n is equal to the sum of the probabilities of these events

$$P(A_1, \text{ or } A_2, \text{ or} \ldots \text{or } A_n) = \sum_{k=1}^{n} P(A_k), \tag{1.3}$$

where $P(A_k)$ is the probability of occurrence of the event A_k.

If the mutually incompatible events $A_1, A_2, \ldots A_n$ constitute a set, one of these events occurs for certain and, consequently,

$$P(A_1, \text{ or } A_2, \text{ or } \ldots \text{ or } A_n) = 1.$$

Considering (1.3) we obtain

$$\sum_{k=1}^{n} P(A_k) = 1, \tag{1.4}$$

i.e., the sum of the probabilities of occurrence of the random events comprising a set is equal to unity.

If a set consists of two events $\underline{A}$ and $\underline{\bar{A}}$, it follows from formula (1.4) that

$$P(A) = 1 - P(\bar{A}). \tag{1.5}$$

Since the two events $\underline{A}$ and $\underline{\bar{A}}$, comprising a set, are opposites, formula(1.5) makes it possible to find the probability of occurrence of event $\underline{A}$ if the probability

of the opposite event $\bar{A}$ is known.

Let all n random events comprising a set be equiprobable, i.e.,

$$P(A_k) = p = \text{const}$$

for any $k = 1, 2, \ldots n$. Then according to formula (1.4) we find the probability p of the occurrence of one of n equiprobable events, which comprise a set, to be

$$\sum_{k=1}^{n} p = np = 1,$$

or

$$p = \frac{1}{n}. \tag{1.6}$$

The probability of realizing one of $m \leqslant n$ events, which enter into a set of n equiprobable events, is equal in accordance with (1.3) and (1.6) to

$$P(A_1, \text{ or } A_2, \text{or} \ldots \text{or } A_m) = \sum_{k=1}^{m} \frac{1}{n} = \frac{m}{n}. \tag{1.7}$$

Suppose, for instance, that firing practice is taking place at a target which is divided into ten zones. Suppose further the region lying outside the target area to be worth zero. Let us denote by A_k a random event consisting of a hit on the k-th zone of the target($k=0,1,2,\ldots 10$) with a single shot. Since with a single shot only one zone of the target is hit, any pair of events A_k and A_r ($k \neq r$) constitutes incompatible random events. The aggregate of all eleven events A_0, A_1, ... A_{10} obviously comprises a set.

Let the shooter at the target hit any of its zones (including the zero zone) with equal probability. Then according to formula (1.6) the probability of hitting any of the zones of the target will be

$$p = \frac{1}{11}.$$

The probability of hitting the bullseye, into which enter five zones, from the sixth to the tenth inclusive, is in accordance with (1.7) equal to

$$p_1 = \frac{5}{11}$$

Let another, more experienced shooter hit the bullseye with a probability of

$$p_2 = \frac{9}{11} \, .$$

Let us ask the question, what is the probability of hitting the bullseye if both shooters fire simultaneously. Let us designate by A_1 the event consisting of a hit on the bullseye by the first shooter, and by A_2 a hit by the second shooter. If the rule of addition is used in finding the probability of a hit on the bullseye, we obtain

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) = p_1 + p_2 = \frac{5}{11} + \frac{9}{11} = \frac{14}{11}.$$

This is an absurd result, because the probability cannot exceed unity. The mistake lies in the violation of a condition of the applicability of the addition rule, since the two events A_1 and A_2, consisting in the fact that each of the shooters hits the bullseye, are compatible. To obtain a correct answer for the question which was asked, it is necessary to make use of the multiplication rule.

3. The Multiplication Rule

The probability of occurrence of two compatible, dependent random events A and B is equal to the product of the probability of one of these events by the relative probability of occurrence of the other, computed under the assumption that the first event has occurred,

$$P(A \text{ and } B) = P(A)\ P_A(B) = P(B)\, P_B(A). \tag{1.8}$$

Probabilities of two kinds enter into formula (1.8): the absolute probability of event A (or else event B) computed regardless of whether the event B (or else event A) dependent on it has occurred, and the conditional probability of event B (or else event A) computed on the assumption that event A (or else event B) occurred. Therefore the absolute probabilities P(A) and P(B) are sometimes called the a priori (i.e., before the trial) probabilities, and the conditional probabilities $P_A(B)$ and $P_B(A)$, the a posteriori (i.e., after the trial) probabilities, by the trial

being meant here the random experiment as a result of which there can be realized the event on which depends the probability of occurrence of another event, dependent upon it.

The most important special case of the rule of multiplication expressed by formula (1.8) is the case when events A and B, although compatible, are independent. In such a case the occurrence of one of the events in no way alters the probability of occurrence of the other, i.e., the a priori and a posteriori probabilities become equal to each other:

$$P_A(B) = P(B), \quad P_B(A) = P(A). \tag{1.9}$$

For independent events the rule of multiplication is thus expressed by the following formula:

$$P(A \text{ and } B) = P(A) \cdot P(B). \tag{1.10}$$

Formula (1.10) is expanded for the case of an arbitrary number of independent events $B_1, B_2, \ldots B_n$:

$$P(B_1 \text{ and } B_2 \text{ and} \ldots \text{and } B_n) = P(B_1) \cdot P(B_2) \ldots P(B_n) = \prod_{k=1}^{n} P(B_k). \tag{1.11}$$

Let $\overline{B}_k$ denote an event converse to the event B_k. Then the occurrence of but one (regardless of which one) of the events B_k excludes the possibility of the simultaneous occurrence of all the events $\overline{B}_1, \overline{B}_2, \ldots \overline{B}_n$. Therefore in accordance with (1.5) $P(B_1, \text{ or } B_2, \text{ or } \ldots \text{ or } B_n) = 1 - P(\overline{B}_1 \text{and} \overline{B}_2 \text{ and} \ldots \text{and} \overline{B}_n)$. Since $\overline{B}_1, \overline{B}_2, \ldots \overline{B}_n$ are mutually independent, then, according to the rule of multiplication

$$P(\overline{B}_1 \text{ and } \overline{B}_2 \text{ and } \ldots \text{and } \overline{B}_n) = \prod_{k=1}^{n} P(\overline{B}_k).$$

Furthermore,

$$P(\overline{B}_k) = 1 - P(B_k).$$

Consequently,

$$P(B_1 \text{ or } B_2 \text{ or} \ldots \text{or } B_n) = 1-\prod_{k=1}^{n}[1-P(B_k)]. \qquad (1.12)$$

Formula (1.12) makes it possible to compute the probability of occurrence of at least one of the compatible, mutually independent events B_1, B_2, ...B_n according to the given probabilities of these events*.

With the aid of formula (1.12) it is possible very simply, for instance, to find the correct answer to the question asked above about the probability of hitting the bullseye in a target if both shooters fire simultaneously. The sought after probability is equal to

$$P(A_1 \text{ or } A_2) = 1 - [1 - P(A_1)]\ [1 - P(A_2)] = \frac{109}{121}$$

Returning to the general formula (1.8), it will be noted that it directly leads to

$$P_A(B)=\frac{P(B)\cdot P_B(A)}{P(A)}. \qquad (1.13)$$

Formula (1.13) makes it possible on the basis of the absolute (a priori) probabilities of two events and relative (a posteriori) probability of one of them to find the relative probability of the other event.

In certain problems it is necessary to determine the probability of event <u>A</u> which occurs with one of <u>n</u> mutually incompatible events B_1, B_2, ...B_n . These <u>n</u> mutually incompatible events are frequently called hypotheses linked with the occurence of event <u>A</u>. Using the rule of addition, let us express the probability of event <u>A</u> in the form of the sum

$$P(A)=P(A \text{ and } B_1, \text{ or } A \text{ and } B_2, \ldots \text{ or } A \text{ and } B_n)=\sum_{k=1}^{n} P(A \text{ and } B_k). \qquad (1.14)$$

* This formula generalizes the addition rule for the case of compatible events. For greater detail see in [3].

Every term of this sum is, in accordance with the multiplication rule, equal to

$$P(A + B_k) = P(B_k)\, P_{B_k}(A)$$

and, consequently

$$P(A) = \sum_{k=1}^{n} P(B_k) P_{B_k}(A). \qquad (1.15)$$

The relationship (1.15) is called the total probability formula. This formula makes it possible to determine the probability of event A, if the a priori probabilities of hypotheses B_1, B_2, ...B_n and the a posteriori probabilities of event A are known, under the condition that one of the hypotheses be confirmed.

Let it now be necessary to find the a posteriori probability of hypothesis B_i under the condition that event A has occurred. For this it is sufficient to make use of formula (1.13)

$$P_A(B_i) = \frac{P(B_i) P_{B_i}(A)}{P(A)}.$$

Utilizing also the total probability formula, we obtain

$$P_A(B_i) = \frac{P(B_i) P_{B_i}(A)}{\sum_{k=1}^{n} P(B_k) P_{B_k}(A)}. \qquad (1.16)$$

The equality (1.16) is called the Bayes formula.

4. Examples of Practical Applications of Fundamental Rules.

As the first important example of the practical application of the basic rules of probability theory let us consider the question of calculating the reliability of operation of a system of several radio-engineering devices. The elements of such a system can be tubes, circuits, or stations.

Specifically we shall examine a radio-relay communications line, consisting of

n stations. A typical problem, which must often be solved in practice, it to find the probability of a breach in communication between the terminal points of the radio line over a definite interval of time T, with the assumption that all stations operate independently of one another.

Let the probability of the breakdown of the k-th station ($k = 1, 2, \ldots n$) over the interval of time T equal p_k. Let us designate by B_k the event of the breakdown within the indicated interval of the k-th station. Since the breakdown of at least one station breaks the communication between the terminal points of the line, to solve the indicated problem it is necessary to find the probability

$$P_{\text{нар}} = P(B_1, \text{ or } B_2 \text{ or } \ldots \text{ or } B_n).$$

The events B_k are compatible, since during T several stations can go out of operation. The use of the addition rule for computing $P_{\text{нар}}$ would not in this case be permissible.

Let us designate by $\bar{B}_k$ an event converse to B_k, i.e., that the k-th station operates without breakdown in the time interval T. The event converse to a break in communication on the line consists in the fact that all stations are operating and, consequently, in accordance with (1.5)

$$P_{\text{нар}} = 1 - P(\bar{B}_1, \text{ and } \bar{B}_2 \text{ and } \ldots \text{ and } \bar{B}_n)$$

It is assumed that all stations operate independently, and therefore, using the multiplication rule, we obtain

$$P(\bar{B}_1 \text{ and } \bar{B}_2 \text{ and } \ldots \text{ and } \bar{B}_n) = P(\bar{B}_1)\, P(\bar{B}_2) \ldots P(\bar{B}_n).$$

The probability of event $\bar{B}_k$, as converse to B_k, is equal to $1 - p_k$. Therefore

$$P(\bar{B}_1 \text{ and } \bar{B}_2 \ldots \text{ and } \bar{B}_n) = (1 - p_1)\,(1 - p_2) \ldots (1 - p_n).$$

The sought probability of a break in communication between the terminal points of the line is equal to

$$P_{\text{нар}} = 1 - (1 - p_1)(1 - p_2) \ldots (1 - p_n). \tag{1.17}$$

Formula (1.17) links the probability of a break in communication on the line with the probabilities of damage to its individual elements - the intermediate stations. This result follows directly from (1.12). In the example treated, the derivation of this formula was repeated for a concrete problem.

If the probability of the breakdown of one station is low, and the total number of stations is small, we obtain from (1.17) the approximate equality

$$P_{\text{нар}} \approx p_1 + p_2 + \ldots + p_n, \tag{1.18}$$

i.e., with low probabilities of damage to individual stations the probability of a break in communication on the line for any interval of time T is approximately equal to the sum of the probabilities of the breakdown of the intermediate stations for the same time T.

In a special case, when the probability of the breakdown of any one of n stations on the line is constant, i.e., $p_k = p = \text{const}$, formula (1.17) takes the form of

$$P_{\text{нар}} = 1 - (1 - p)^n. \tag{1.19}$$

When $p \ll \frac{1}{n}$ from (1.19) we find

$$P_{\text{нар}} = 1 - \left(1 - \frac{np}{n}\right)^n \approx 1 - e^{-np} \approx np. \tag{1.20}$$

It can be seen from formula (1.20) that, with a large number of intermediate stations, in order to provide for sufficiently high operational reliability of the entire line it is necessary to have rigid specifications for the reliable operation of each of the stations. Thus, for instance, when the number of stations on the line is $n = 100$, and the permissible probability of communication interruption is $P_{\text{нар}} = 0.01$ the probability of breakdown of each station must be equal to $p = 0.0001$.

It is perfectly evident that the formulas cited above can be utilized for computing the reliability of operation of a station consisting of several units, or for computing the reliability of operation of a unit, damage to one element of which

would lead to interrupted operation of the entire unit.

To raise the reliability of operation of systems consisting of a large number of elements, a standby system can be incorporated wherein several elements operate in parallel. It is clear that if the probability of one element going out of operation is equal to p_1, the probability of the simultaneous breakdown of n independent parallel elements, whose operational unreliabilities are defined respectively by the probabilities $p_1, p_2, \ldots p_n$, is equal to the product of $p_1 \cdot p_2 \cdot \ldots \cdot p_n$, i.e., is considerably less than p_1.

The following example illustrates the application of the Bayes formula in problems dealing with the transmission of a telegraphic signal with the presence of noise [7].

Let us assume that for the transmission of two messages - "yes" or "no" - a simple telegraph system is used and that the messages are transmitted by two different signals, distinguishable in the receiver by the lighting of a red and a green lamp. If the signal were always transmitted without distortion it would be possible on the basis of the signal received (for instance on the basis of the lighting of the green lamp) to give an unambiguous answer to the question of which message was sent (for instance, "yes"), i.e., the a posteriori probability of "yes" under the condition that the green lamp is lit would be equal to one.

Due to noise distortion the signal received will not always reliably indicate which message was transmitted, i.e., there will be cases when the green lamp lights up at the transmission of the message "no", and the red lamp at the transmission of "yes".

The question is asked, what is the probability that the operator decoding the message (i.e., determining the message from the signal received) does not make a mistake in asserting, with the lighting of the green lamp, that the message transmitted was "yes"? In other words, it is necessary to determine the a posteriori probability of the fact that the message "yes" was transmitted under the condition that the green lamp was lit.

In connection with event ($\underline{A}$) - the lighting of the green lamp - two hypotheses can be expressed: ($\underline{B}_1$) - the message "yes" was transmitted and ($\underline{B}_2$) - the message "no" was transmitted.

To determine the a posteriori probability of the fact that the message "yes" was transmitted with the lighting of the green lamp, it is necessary to know in advance what part of the total number of messages consists of "yes" messages, i.e., the a priori probability of the transmission of the "yes" message. In addition, it is necessary to know the statistical properties of noise distorting the transmission, i.e., the a posteriori probabilities of the lighting of the green lamp with the transmission of the signal "yes" and of the signal "no".

Let us assume that among the messages transmitted, "yes" and "no" are encountered in a ratio of 5 : 3. Then the a priori probability of the transmission of the "yes" message is equal to

$$P(\text{"yes"}) = \frac{5}{8}$$

and the a priori probability of the converse event, i.e., the transmission of the "no" message, is equal to

$$P(\text{"no"}) = \frac{3}{8}$$

Let us further assume that the statistical properties of noise are such that 2/5 of the "yes" messages transmitted are distorted, and 1/3 of the "no" messages. Then the a posteriori probability of the lighting of the green lamp with the transmission of the "yes" message (undistorted transmission) is equal to

$$P_{\text{"yes"}}(\text{"green"}) = \frac{3}{5}$$

and the a posteriori probability of the lighting of the green lamp with the transmission of the "no" message (distorted transmission) is equal to

$$P_{\text{"no"}}(\text{"green"}) = \frac{1}{3}$$

The desired a posteriori probability of the transmission of "yes", under the

condition that the green lamp lights up, can now be determined by means of formula (1.13), which follows directly from the multiplication rule,

$$P_{\text{"green"}}(\text{"yes"}) = \frac{P(\text{"yes"}) \cdot P_{\text{"yes"}}(\text{"green"})}{P(\text{"green"})} \qquad (1.21)$$

The quantities in the numerator are determined by the statistical properties of the messages and noise and for the example under consideration are indicated above. The a priori probability of whether the green lamp will light up is not known in advance, but it can be computed by the formula of total probability. Actually, the lighting of the green lamp may indicate two incompatible hypotheses: either the "yes" message was transmitted, or the "no". Then according to (1.14)

$$P(\text{"green"}) = P(\text{"yes"})\, P_{\text{"yes"}}(\text{"green"}) + P(\text{"no"})\, P_{\text{"no"}}(\text{"green"})$$

or, substituting the given values of the probabilities which characterize the messages and the noise, we obtain the a priori probability of the lighting of the green lamp

$$P(\text{"green"}) = \frac{5}{8} \cdot \frac{3}{5} + \frac{3}{8} \cdot \frac{1}{3} = \frac{1}{2}$$

Although the "yes" messages are transmitted more frequently, it turns out that due to the presence of noise the lighting-up of the green or the red lamp is equiprobable. Now it is not difficult according to (1.21) to find the desired probability, i.e, the probability of the correctness of the answer "yes" when the green lamp is lit

$$P_{\text{"green"}}(\text{"yes"}) = \frac{5/8 \cdot 3/5}{1/2} = \frac{3}{4}$$

The same result is obtained directly from (1.16), i.e, from the Bayes formula, without the preliminary computation of the a priori probability of the lighting of the green lamp.

In the same manner, formula (1.16) can be used to find that with the lighting of the red lamp the a posteriori probability that the "no" message was transmitted

is equal to

$$P_{\text{"red"}}(\text{"no"}) = \frac{P(\text{"no"})\, P_{\text{"no"}}(\text{"red"})}{P(\text{"no"})\, P_{\text{"no"}}(\text{"red"}) + P(\text{"yes"})\, P_{\text{"yes"}}(\text{"red"})} \tag{1.22}$$

or

$$P_{\text{"red"}}(\text{"no"}) = \frac{3/8 \cdot 2/3}{3/8 \cdot 2/3 + 5/8 \cdot 2/5} = \frac{1}{4} : \frac{1}{2} = \frac{1}{2}$$

5. Series of Independent Tests

Numerous problems, the solutions to which have widespread practical applications, fit the following system of a serios of independent tests.*

Let there be made n independent tests, the probability of occurrence of random event A being the same for each test and equal to p. The probability of the non-occurrence of event A in the test, i.e., the probability of the converse event, is equal to $q = 1-p$. It is required to find the probability that after n tests the event A will occur exactly k times $(k \gg n)$. The solution to this problem is obtained by simple employment of the addition and multiplication rules.

With respect to the order of occurrence of event A exactly k times, there can be expressed a number of mutually exclusive hypotheses. One of these hypotheses (B_1) consists in the fact that event A occurs in the first k experiments and does not take place in the n - k succeeding ones. Then according to formula (1.11) we find

$$P_n(k \text{ and } B_1) = \underbrace{p \cdot p \ldots p}_{\text{k times}} \underbrace{q \cdot q \ldots q}_{(n-k) \text{ times}} = p^k q^{n-k}.$$

Another hypothesis (B_2) consists in the fact that event A does not occur in the first test, then takes place k times in a row and does not appear again in the remaining n - k - 1 experiments. In such a case

$$P_n(k \text{ and } B_2) = q \cdot \underbrace{p \cdot p \ldots p}_{\text{k times}} \underbrace{q \cdot q \ldots q}_{(n-k-1) \text{ times}} = p^k q^{n-k}$$

*The term "test" is fully equivalent to the previously introduced term "random experiment" and is employed here in accordance with the tradition established in the literature.

It is perfectly obvious that for any hypothesis B_S on the order of occurrence of event A exactly k times and its non-occurrence $n - k$ times in n independent tests, the probability P_n (k and B_S), computed according to formula (1.11), will after the appropriate transposition of the co-multiples be equal to $p^k q^{n-k}$.

Thus any order of the occurrence of event A exactly k times with n independent tests turns out to be equiprobable. The number of possible hypotheses on the order of the occurrence of event A exactly k times with n tests is equal to the number of combinations from n elements to k. This value is usually denoted by the symbols C_n^k or $\binom{n}{k}$ and is expressed in the form of the ratio of factorials

$$C_n^k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Since the hypotheses B_1, B_2, ...$B_{\binom{n}{k}}$ are mutually exclusive, by employing the total probability formula (1.14) we obtain the following expression for the desired probability $P_n(k)$ of the occurrence of event A exactly k times with n independent experiments

$$P_n(k) = \sum_{s=1}^{\binom{n}{k}} P_n(k \text{ and } B_s) = \binom{n}{k} p^k q^{n-k}. \qquad (1.23)$$

It is not difficult to note that $P_n(k)$ is equal to the coefficient of x^k in the expansion of the binomial $(q + px)^n$ by powers of x. For this reason formula (1.23) is frequently called the binomial formula.

All the possible quantities of the occurrence of event A in n tests (k=0, 1, ..., n) constitute a set of mutually exclusive events, since in the tests event A either does not occur at all, or occurs some definite number of times. Therefore in accordance with (1.4)

$$\sum_{k=0}^{n} P_n(k) = 1. \qquad (1.24)$$

Formula (1.24) follows also from the expansion of the binomial $(q + px)^n$, from which when $\underline{x} = 1$ is obtained

$$(q+p)^n = 1^n = 1 = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$$

The simplest example of a system of a series of independent tests is the probl examined in Section 4 of the reliability of a system of components with the condit that the reliability of any of the components is the same and equal to $\underline{q} = 1 - \underline{p}$. Each of the independent experiments consists of testing, whether one of the components of the system has gone out of commission (event $\underline{A}$). The probability that event $\underline{A}$ will not take place at all ($\underline{k} = 0$) is, in accordance with (1.23) equal to $(1 - p)^n$, wherefrom the probability of interrupted operation will be

$$P_{\text{нар}} = 1 - (1 - p)^n,$$

which does not differ from (1.19).

Of greater interest in practice is not the probability of the occurrence of an event a definite number of times, but the probability that the number of times the event occurs with $\underline{n}$ independent experiments lies within the definite limits of from $\underline{k}_1$ to $\underline{k}_2$.

Employing the addition rule, we find this probability to be

$$P_n(k_1 \leqslant k \leqslant k_2) = P_n(k_1) + P_n(k_1 + 1) + \ldots + P_n(k_2),$$

or, employing (1.23)

$$P_n(k_1 \leqslant k \leqslant k_2) = \sum_{k=k_1}^{k=k_2} \binom{n}{k} p^k q^{n-k}. \tag{1.25}$$

6. Problem of the Number of Channels in a Communication Line.

The following problem of the necessary number of channels in a communication line can serve as one of the numerous examples of a system of a series of independent experiments.

Let a communication line connect point $\underline{M}$ with 10 subscribers at point $\underline{N}$. In order to gain a concept of the expected load on this line, let us assume that each of these subscribers uses the line for an average of 12 minutes per hour. Then the a priori probability that the line will be required is equal to $\underline{p} = 1/5$. It can also within certain tolerances be considered that the calls of any two subscribers are independent. What is the probability that the line will be required simultaneously by $\underline{k}$ subscribers ($\underline{k}$ = 1, 2, ... 10)?

The solution of this problem is provided by the binomial formula (1.23). Here the event $\underline{A}$ is a call to point $\underline{M}$ from point $\underline{N}$. The series of experiments consists in testing, did or did not each of the 10 subscribers make a call.

The probability that the line will be required by only one subscriber is equal to

$$P_{10}(1)=\binom{10}{1}\cdot\frac{1}{5}\cdot\left(\frac{4}{5}\right)^{9}=0{,}268.$$

The probability that the line will be required simultaneously by two subscribers is equal to

$$P_{10}(2)=\binom{10}{2}\cdot\left(\frac{1}{5}\right)^{2}\cdot\left(\frac{4}{5}\right)^{8}=0{,}302.$$

The results of the subsequent calculations are presented in the table immediately following.

Table 1

k	3	4	5	6	7	8	9	10
$P_{10}(k)$	0,201	0,088	0,026	0,0055	0,0008	0,00007	0,000004	0,0000001

Having also determined the probability that the line will not be required by any of the subscribers,

$$P_{10}(0)=\binom{10}{0}\left(\frac{1}{5}\right)^{0}\left(\frac{4}{5}\right)^{10}=0{,}108,$$

it is not difficult to see that the sum of all the computed probabilities is, in accordance with (1.24), equal to unity.

Above were computed the probabilities that the line would be required simultaneously by exactly k subscribers. Let us now determine the probabilities that the line would be required simultaneously by not more than k subscribers (k = 1, 2, ... 10), for which we utilize formula (1.25). Then we obtain

$$\begin{aligned}
P_{10}(0 \leqslant k \leqslant 1) &= 0{,}376,\\
P_{10}(0 \leqslant k \leqslant 2) &= 0{,}678,\\
P_{10}(0 \leqslant k \leqslant 3) &= 0{,}879,\\
P_{10}(0 \leqslant k \leqslant 4) &= 0{,}967,\\
P_{10}(0 \leqslant k \leqslant 5) &= 0{,}993,\\
P_{10}(0 \leqslant k \leqslant 6) &= 0{,}9985.
\end{aligned}$$

Subsequent computations will give probability values differing from unity by no more than in the fourth decimal place. An analysis of the values obtained shows that it would be pointless in the case under consideration to have 10 channels for communication between point N and point M, since the probability that a connection will be required simultaneously by more than 5 subscribers

$$P_{10}(6 \leqslant k \leqslant 10) = 1 - P_{10}(0 \leqslant k \leqslant 5) = 1 - 0{,}993 = 0{,}007$$

is very small and may in practice be disregarded. This result shows that for practically trouble-free service to 10 subscribers (on the condition that each of them uses the line for an average of 12 minutes every hour) it is sufficient to have but 5 channels on the line. With such a quantity only seven calls out of a thousand could not be immediately provided with a free channel.

Figure 1 shows a graph of $p_{10}(k)$ as a function of k when p = 1/5, based on the results of the above computations. This function is determined only for the integer values of the argument, but in Figure 1 these points are arbitrarily joined by a

broken line. It can be seen from this graph that the function $P_{10}(k)$ reaches a maximum when $k = 2$, i.e., the most probable event is that the line be required simultaneously by two subscribers.

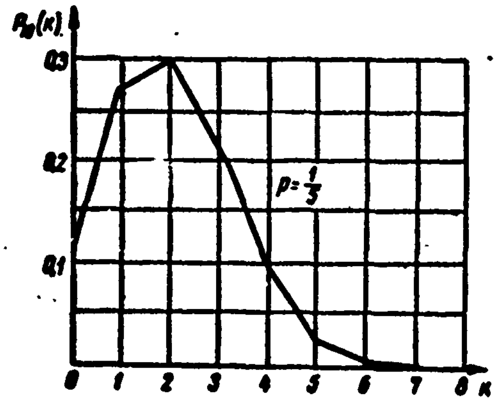


Fig. 1. Probabilities of $P_{10}(k)$ with $p = 1/5$.

The presence of a probability maximum, defined by the binomial formula (1.23), is characteristic also of the general case, the value of the most probable number, k_0, of occurrences of an event depending both on the number, n, of independent tests, and on the probability p of the occurrence of the event at each test.

To determine this most probable number k_0 in the general case it is sufficient to examine the relation

$$\frac{P_n(k)}{P_n(k-1)} = \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k-1} p^{k-1} q^{n-k+1}} = \frac{(n-k+1)p}{kq} = 1 + \frac{(n+1)p-k}{kq}. \tag{1.26}$$

If $k < (n+1)p$, then $P_n(k) > P_n(k-1)$, and when $k > (n+1)p$, $P_n(k) < P_n(k-1)$. Consequently, the most probable number, k_0, of occurrences of an event, at which the probability $P_n(k)$ reaches a maximum, must be equal to the integer contained in the value of $(n+1)p$.

$$k_0 = [(n+1)p]. \tag{1.27}$$

Here the symbol $[x]$ signifies the integer of the number x.

In the example under discussion $n = 10$, $p = 1/5$ and consequently, $k_0 = \left[\frac{11}{5}\right] = 2$.

If $(n + 1)p$ is an integer, then as follows from (1.25), there exist two maximum probabilities which are equal to each other

$$P_n(k_0) = P_n(k_0 - 1).$$

7. The Asymptotic Formulas of Laplace

In cases where the number of independent tests is large, the direct computation of probabilities according to formulas (1.23) and (1.25) raises great difficulties, since the necessary determination of the binomial coefficients is linked with the computation of factorials with large arguments. Thus, for instance, if in the problem examined above the number of subscribers at point N were not 10 but 100, for its solution it would have been necessary to compute the values of (100)! (99)! ... The value of the factorial may be obtained with sufficient accuracy by the use of the so-called asymptotic formula of Stirling.

This asymptotic formula has the form*

$$m! \sim \sqrt{2\pi m} \cdot m^m \cdot e^{-m}. \tag{1.28}$$

Table 2 presents a comparison of the approximate values of factorials computed by formula (1.28) with the exact values.

Table 2

m	$m!$ (exact)	$m!$ by (1.28)	Relative error in %
1	1	0,922	8
2	2	1,919	4
5	120	118,019	2
10	$3{,}6288 \cdot 10^6$	$3{,}5986 \cdot 10^6$	0,8
100	$9{,}3326 \cdot 10^{157}$	$9{,}3249 \cdot 10^{157}$	0,08

* The symbol $\sim$ (asymptotic equality) signifies that the ratio of the two expressions related by this symbol tends toward unity with the unlimited growth of m.

We transform the binomial formula (1.23) by means of Stirling's formula

$$P_n(k)=\frac{n!}{k!\,(n-k)!}p^k q^{n-k}\sim$$

$$\sim\frac{\sqrt{2\pi n}\,n^n e^{-n}p^k q^{n-k}}{\sqrt{2\pi k}\,k^k e^{-k}\sqrt{2\pi(n-k)}\,(n-k)^{n-k}e^{-(n-k)}},$$

wherefrom

$$P_n(k)\sim\sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}. \tag{1.29}$$

But even with the use of the approximate formula (1.29) the computation of the probabilities $P_n(k)$ and particularly of sums like $\sum_{k=k_1}^{k=k_2} P_n(k)$ with a large n remains sufficiently labor-consuming.

However, under certain additional conditions from formula (1.29), there can be obtained a very simple asymptotic formula which makes it possible to determine the indicated probabilities with a sufficient degree of accuracy.

Let us designate the deviation of the number of occurrences k of an event with n independent tests from its most probable value of k_0 by

$$\delta_k=k-k_0=k-[(n+1)p].$$

In the case of a large n the value of $(n+1)p$ differs relatively little from its integer, and therefore when $n \gg 1$ it is possible to assume that $\delta_k = k - np$.

Then

$$k=np+\delta_k,\quad n-k=nq-\delta_k. \tag{1.30}$$

Substituting (1.30) into (1.29) and designating by σ the value

$$\sigma=\sqrt{npq}, \tag{1.30'}$$

we obtain

$$P_n(k) \sim \frac{1}{\sqrt{2\pi\sigma^2}}\left(1+\frac{\delta_k}{np}\right)^{-\left(np+\delta_k+\frac{1}{2}\right)}\left(1-\frac{\delta_k}{nq}\right)^{-\left(nq-\delta_k+\frac{1}{2}\right)}$$

or

$$-\ln[\sqrt{2\pi\sigma^2}P_n(k)] \sim \left(np+\delta_k+\frac{1}{2}\right)\ln\left(1+\frac{\delta_k}{np}\right)+$$
$$+\left(nq-\delta_k+\frac{1}{2}\right)\ln\left(1-\frac{\delta_k}{nq}\right).$$

Since $\left|\delta_K\right| \leqslant np$ and $\left|\delta_K\right| < nq$ $(k \neq n)$, by expanding the logarithms into exponential series and by limiting ourselves to the first three terms of the expansion, we obtain after simple transformations

$$-\ln[\sqrt{2\pi\sigma^2}P_n(k)] \sim \frac{\delta_k^2}{2\sigma^2}-\frac{\delta_k^3}{6\sigma^4}(1-2p)+O\left(\frac{\delta_k^4}{\sigma^6}\right), \tag{1.31}$$

where $O\left(\frac{\delta_K^4}{\sigma^6}\right)$ is the sum of all the terms of the expansion in which the order of smallness of each item is not lower than $\frac{\delta_K^4}{\sigma^6}$.

Let us now examine only that range of the values of k for which $\frac{\delta_K^3}{\sigma^4} \longrightarrow 0$ with an unlimited increase in n. Then from (1.31) there follows the asymptotic equality

$$P_n(k) \sim \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\delta_k^2}{2\sigma^2}}. \tag{1.32}$$

Formula (1.32), which is sometimes called the Laplace formula, is the desired asymptotic approximation to the probability $P_n(k)$, the exact value of which is pro-

vided by the binomial formula (1.23). It will be the more precise, the better is fulfilled the condition

$$\frac{|\delta_k^3|}{\sigma^4} \ll 1. \tag{1.33}$$

Let us note that, for constant k and n, formula (1.32) provides an approximation with maximum precision when p = q = ½. In such a case, the errors become small values of the order of $\frac{\delta_K{}^4}{\sigma^6}$.

It follows from (1.32) that the probability of the most probable number of occurrences $k = k_0$ of an event will be

$$P_n(k_0) \sim \frac{1}{\sqrt{2\pi\sigma^2}}. \tag{1.32'}$$

Making use of (1.32), it is not difficult to write an expression of the probability that the number of occurrences of an event will be contained in the limits of from k_1 to k_2:

$$P_n(k_1 \leqslant k \leqslant k_2) = P_n\left(\frac{k_1 - k_0}{\sigma} \leqslant \frac{\delta_k}{\sigma} \leqslant \frac{k_2 - k_0}{\sigma}\right) \sim$$
$$\sim \sum_{k=k_1}^{k=k_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\delta_k^2}{2\sigma^2}}.$$

Having noted that

$$\frac{1}{\sigma} = \frac{\delta_{k+1} - \delta_k}{\sigma} = \frac{\Delta\delta_k}{\sigma},$$

we rewrite the last equality in the form

$$P_n\left(a \leqslant \frac{\delta_k}{\sigma} \leqslant b\right) \sim \frac{1}{\sqrt{2\pi}} \sum_{k=k_1}^{k=k_2} \frac{\Delta\delta_k}{\sigma} e^{-\frac{\delta_k^2}{2\sigma^2}}, \tag{1.34}$$

where are assumed the designations

$$a=\frac{k_1-k_n}{\sigma},\ b=\frac{k_2-k_n}{\sigma}. \tag{1.35}$$

The summation in the right part of (1.34) may be regarded as an integral one, which with a sufficiently large n will differ little from the integral

$$\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{z^2}{2}}dz.$$

Thus the probability that the number of occurrences of an event with n independent tests falls within the limits of from k_1 to k_2 may be computed by means of the asymptotic formula

$$P_n(k_1\leqslant k\leqslant k_2)=P_n\left(a\leqslant\frac{\delta_k}{\sigma}\leqslant b\right)\sim\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{z^2}{2}}\,dz, \tag{1.36}$$

where the values of a and b are defined in (1.35).

Let us note that with a large number of tests formula (1.36) is applicable also in those cases when condition (1.33) is violated for borderline values of k, close to k_1 and k_2, if $k_1 - k_2$ is large. This is because, with a large n, the probability of the occurrence of an event a precisely defined number of times is extremely small [cf (1.32)], and even large errors in the determination of individual probabilities for borderline values of k cannot substantially influence the overall value of P_n $(k_1 \leqslant k \leqslant k_2)$.

Formula (1.36) is the analytic expression of the so-called integral theorem of Laplace. Assuming the designation

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{z^2}{2}}\,dz, \tag{1.37}$$

it is possible to rewrite (1.36) as

$$P_n\left(a\leqslant\frac{\delta_k}{\sigma}\leqslant b\right)\sim F(b)-F(a). \tag{1.38}$$

In Appendix II there is presented a table of the values of $F(x)$ and of the integrand function

$$w(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

The table is prepared only for positive arguments, since the function $w(x)$ is even, and the values of $F(-x)$ are obtained from the relationship

$$F(-x)=1-F(x). \tag{1.39}$$

Often in place of the Laplace function there is investigated and tabulated the so-called integral of probability (function of errors, Kramp function)

$$\Phi(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}dz=2F(x\sqrt{2})-1 \tag{1.40}$$

and the function

$$F_0(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{z^2}{2}}dz=F(x)-\frac{1}{2}, \tag{1.41}$$

for which the relationship (1.39) is replaced by the more simple

$$F_0(-x)=F_0(x).$$

The function $F(x)$ plays an extremely important role in probability theory. Its properties will be presented in greater detail in the following chapter.

For the solution of certain problems it is convenient to make use of the following special cases of the general formula (1.38)

$$P_n\left(\frac{\delta_k}{\epsilon}\leq b\right)\sim F(b), \tag{1.42}$$

$$P_n\left(\frac{\delta_k}{\epsilon}\geq a\right)\sim 1-F(a), \tag{1.43}$$

$$P_n\left(\frac{|\delta_k|}{\epsilon}\leq \alpha\right)\sim 2F(\alpha)-1=\Phi\left(\frac{\alpha}{\sqrt{2}}\right). \tag{1.44}$$

An important consequence arises out of formula (1.44). If $\alpha = 3$, then

$$P_n\left(\frac{|\delta_k|}{\sigma} < 3\right) = P_n(|k - k_0| \leqslant 3\sigma) = 2F(3) - 1 = 0{,}997. \tag{1.45}$$

An event whose probability of occurrence is equal to 0.997 is customarily considered as being practically certain. Thus it is practically certain that with a large number of independent experiments the deviations of the number of occurrences of an event from the most probable will lie within the limits of $|k - k_0| \leqslant 3\sigma$.

As an illustration of the application of the Laplace asymptotic formulas (1.32) and (1.38) let us return to the problem discussed in the previous section for the case when the number of subscribers at point N grows from ten to a thousand. The solution of the problem in this case by means of the binomial formula (1.23) is extremely cumbersome, whereas the Laplace asymptotic formulas make it possible to perform the necessary calculations very simply.

Let each of the thousand subscribers engage a line an average of six minutes per hour. The a priori probability that the line will be required equals $p = 1/10$. The most probable number of calls equals in this case

$$k_0 = np = \frac{1000}{10} = 100.$$

However, the probability that exactly 100 calls will take place is in this case extremely small. Actually, it follows from (1.32') that

$$P_n(k_0) = \frac{1}{\sqrt{2\pi\sigma^2}} = \frac{1}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{6{,}28 \cdot 100 \cdot 0{,}9}} \sim 0{,}04.$$

If the line has 100 channels, the probability that all channels will be occupied and that at least one of the subscribers will get a "busy" signal is, in accordance with (1.43) equal to

$$P_{1000}(k > 100) = P_{1000}\left(\frac{\delta_k}{\sigma} > 0\right) = 1 - F(0) = 0{,}5.$$

It can be considered practically certain that the number of simultaneous calls

will not exceed the value of*

$$k \leq k_0 + 3\sigma = 100 + 3\sqrt{100 \cdot 0{,}9} = 128{,}5.$$

This result shows that for the trouble-free servicing of 1,000 subscribers, under the condition that each of them talks on the average of six minutes each hour, the communication line should have only 130 channels.

It is possible to solve the converse problem: to how many subscribers can the line provide practically trouble-free service, if with the same load (i.e., when $p = 1/10$) the number of channels is increased to 200? The desired number of subscribers is determined from the equation

$$k_0 + 3\sigma - 200 = 0,$$

which, considering (1.27) and (1.30'), is reduced to the quadratic equation with respect to σ

$$\frac{10}{9}\sigma^2 + 3\sigma - 200 = 0,$$

wherefrom

$$\sigma \approx 12 \text{ и } n = \frac{\sigma^2}{pq} = 144 \cdot \frac{100}{9} = 1600.$$

It is also not difficult to provide an answer to another question of practical interest: will the total number of channels decrease if the communication line between points M and N were split up into several lines, each of them serving its respective part of the subscribers of point M. For instance, if the line examined above were split up into two, each serving 500 subscribers, the necessary number of channels for practically trouble-free service on each line will amount to

$$k = 50 + 3\sqrt{50 \cdot 0{,}9} \approx 70.$$

The total number of channels in the two lines will be equal to 140, i.e., to 10 channels more than with one communication line.

* The exact value of the probability of the inequality cited below is 0.998.

8. Asymptotic Formula of Poisson

In many practical problems pertaining to a system of a large number of independent tests ($n \gg 1$), the probability of occurrence of an event with a single test is relatively low, so that

$$p = \frac{\lambda}{n}, \tag{1.46}$$

where λ is a constant positive value.

In this case the asymptotic approximation (1.32) for the probability $P_n(k)$ yields considerable errors, and it is necessary to use a different asymptotic formula, arrived at by Poisson.

Let us investigate the probability that with n tests the event will not occur at all. On the basis of (1.23) and (1.46) this probability may be presented in the following form:

$$P_n(0) = (1-p)^n = \left(1 - \frac{\lambda}{n}\right)^n,$$

wherefrom

$$\ln P_n(0) = n \ln\left(1 - \frac{\lambda}{n}\right) = -\lambda - \frac{\lambda^2}{2n} - \dots \tag{1.47}$$

If

$$\frac{\lambda^2}{n} \ll 1 \quad \text{or} \quad p \ll \frac{1}{\sqrt{n}}, \tag{1.48}$$

then in the expension of (1.47), it is possible to limit one's self to the first term; then

$$P_n(0) \sim e^{-\lambda}.$$

Analogously

$$P_n(1)=np(1-p)^{n-1}=\frac{np}{1-p}(1-p)^n=$$
$$=\frac{\lambda}{1-\frac{\lambda}{n}}P_n(0)\approx\lambda e^{-\lambda},$$

and in the general case

$$P_n(k)=\frac{n(n-1)\ldots(n-k+1)}{k!}p^k(1-p)^{n-k}=$$
$$=\frac{n(n-1)\ldots(n-k+1)}{n^k\left(1-\frac{\lambda}{n}\right)^k}\frac{\lambda^k}{k!}P_n(0)\sim\frac{\lambda^k}{k!}e^{-\lambda}.$$

The asymptotic formula

$$P_n(k)\sim\frac{\lambda^k}{k!}e^{-\lambda} \tag{1.49}$$

is called the Poisson formula.

Formula (1.20) obtained in Section 4 follows directly from (1.49) when $k = 0$, if it is assumed that the number n of stations is large, and the probability p of the going out of service of one station is small.

A table of values of the Poisson approximation (1.49), for the binomial formula with several values of λ and k, is presented in Appendix I.

Making use of (1.49), let us write an expression for the probability of the occurrence of an event not more than m times:

$$P_n(k\leqslant m)=\sum_{k=0}^{m}P_n(k)\approx\sum_{k=0}^{m}\frac{\lambda^k}{k!}e^{-\lambda}.$$

The last sum may be presented as an integral

$$e^{-\lambda}\sum_{k=0}^{m}\frac{\lambda^k}{k!}=\frac{1}{m!}\int_{\lambda}^{\infty}z^m e^{-z}dz. \tag{1.50}$$

suggesting an integral presentation of the gamma-function $\Gamma(m + 1)$, which for integral values of the argument coincides with the factorial

$$\Gamma(m+1)=m!=\int_0^\infty z^m e^{-z}dz. \tag{1.51}$$

When $\lambda = 0$, the integral in (1.50) does not differ from (1.51).

The integral

$$\Gamma(m+1,\lambda)=\int_0^\lambda z^m e^{-z}dz=\Gamma(m+1)-\int_\lambda^\infty z^m e^{-z}dz \tag{1.52}$$

is called the incomplete gamma-function. By means of this function the probability of the occurrence of an event not more than m times may be presented in the following manner:

$$P_n(k\leqslant m)=1-\frac{\Gamma(m+1,\lambda)}{\Gamma(m+1)}. \tag{1.53}$$

There exist detailed tables of the incomplete gamma-function [4] for a wide range of variation of its arguments. A table of values of the probabilities $p_n(k \leqslant m)$, obtained with the aid of said tables, is presented in Appendix I.

In several applications of the Poisson formula the parameter λ is much greater than unity, although it satisfies the relationship (1.48). In such cases the Poisson approximation (1.49) coincides with that of Laplace (1.32).

Actually, from (1.49) there follows

$$\ln P_n(k)\sim k\ln\lambda-\lambda-\ln k!;$$

using the Stirling formula for $\ln k!$, we obtain

$$\ln P_n(k)\sim k\ln\lambda-\lambda-\frac{1}{2}\ln 2\pi-\left(k+\frac{1}{2}\right)\ln k+k=$$
$$=k-\lambda-k\ln\frac{k}{\lambda}-\frac{1}{2}\ln 2\pi k.$$

Designating $k - \lambda = \delta_K$, let us rewrite the last equality in the form of

$$\ln P_n(k) \sim \delta_k - \left(\delta_k + \lambda + \frac{1}{2}\right) \ln\left(1 + \frac{\delta_k}{\lambda}\right) - \frac{1}{2} \ln(2\pi\lambda).$$

If it is now assumed that $\frac{\delta_K}{\lambda} \ll 1$ (i.e., those values of k are being examined, which are relatively close to the most probable value of λ), then expanding $\ln(1 + \frac{\delta_K}{\lambda})$ in a series of powers of $\frac{\delta_K}{\lambda}$ and limiting ourselves to the first term of the expansion, we obtain

$$\ln P_n(k) \sim -\frac{\delta_k^2}{2\lambda} - \frac{1}{2} \ln(2\pi\lambda),$$

wherefrom

$$P_n(k) \sim \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{\delta_k^2}{2\lambda}},$$

which coincides with the Laplace form of approximation for the binomial formula.

9. The Shot Effect

The number of electrons, emitted by the cathode and reaching the anode in equal intervals of time, fluctuates about a certain mean value. As a result of this phenomenon, which is called the shot effect, there take place random voltage fluctuations in the diode load. It is asked, what is the probability that, during a certain interval of time t, exactly k electrons reach the anode? The answer to this question is provided by the Poisson formula.

Let us assume that the intensity of the electron stream, i.e., the average number ν of electrons reaching the anode per unit of time, is known. Over a sufficiently large interval of time T, the number of emitted electrons will equal $n = \nu T$. These n electrons emerge in the indicated time interval T at random and independently of each other. What interests us is the probability that k of n electrons will emerge in a fixed interval t, which constitutes a part of T.

The emergence of n electrons may be regarded as n consecutive tests. The event A, which either occurs or does not occur as a result of each test, is that a given

electron appears in the time interval t. The probability that any of the electrons appears in the time interval t, i.e., the probability of the occurrence of event A with each test, is equal to $p = \frac{t}{T} = \frac{\nu t}{n}$.

Thus we return again to the system of a succession of independent tests. A peculiarity of the problem in question is the large number of independent tests n and the low probability of the occurrence of an event with a single test. Then, according to the Poisson formula (1.49), substituting $\lambda = np = \nu t$, we find the probability that during the time interval t exactly k electrons reach the anode

$$P_n(k) = \frac{(\nu t)^k}{k!} e^{-\nu t}. \qquad (1.54)$$

There is another way of solving the problem at hand. Preserving the condition of the independence of electron departure, let us subdivide the time interval t into N sections, with a duration of $\Delta t = \frac{t}{N}$ each, and assume that the probability of departure of one electron during the time Δt, equals $\nu \Delta t$. The probability of departure of two or more electrons is an extremely small, second-order one with respect to Δt. The probability that an electron will not reach the anode in the first interval Δt is equal to $1-\nu\Delta t$, and the probability $P_N(0)$ that an electron will not reach the anode in any one of N intervals by virtue of its independence is equal to the product of N co-multipliers of $(1-\nu\Delta t)$, i.e.,

$$P_N(0) = (1-\nu\Delta t)^N = (1-\nu\Delta t)^{\frac{t}{\Delta t}}.$$

Letting now $N \longrightarrow \infty$ or $\Delta t \longrightarrow 0$, we obtain

$$P(0) = e^{-\nu t},$$

i.e., formula (1.54) with $k = 0$. In an analogous manner it is possible to obtain expressions for $P(1)$, $P(2)$, and in the general case for $P(k)$, which leads to the Poisson formula.

The problem of the number of electrons reaching the anode is one of a broad class of problems, in which there is determined the probability of the occurrence of an event a definite number of times during a certain time interval t. To this class belong problems concerning the number of telephone calls made over various

intervals of time, or concerning the number of disintegrated radioactive atoms.

The Poisson formula is also used for solving the problem of the number of sign changes in the period $\underline{t}$ in a generalized telegraph signal (Fig. 2), which is characterized by the fact that the current $\underline{i}(\underline{t})$ may take only two values, $+\underline{h}$ and $-\underline{h}$, but that the instants of sign change are random.* If ν is the average number of sign changes in a unit of time and if the probability of a sign change in an interval $\underline{t}$ (equal to $\gamma \Delta \underline{t}$) does not depend on the pattern of sign changes outside this interval, for the computation of the probability that during the time $\underline{t}$ there will occur exactly $\underline{k}$ sign changes, by a reasoning analogous to that carried out above, we obtain formula (1.54).

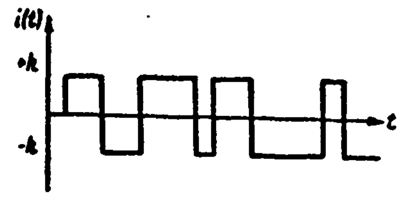


Fig. 2. Generalized telegraph signal

10. Generalization of the System of Independent Tests. The Concept of the Markov Chain

There has up to now been examined a succession of independent events which was characterized by the occurrence or non-occurrence of a definite event $\underline{A}$ with each test. In certain cases the outcome of each test can be not one of the two events $\underline{A}$ or $\bar{A}$, but a set of $\underline{m}$ incompatible events $A_1, A_2, \ldots A_m$. Let us assume that the probability of occurrence of event A_i with each test is equal to p_i ($\underline{i} = 1,2,\ldots \underline{m}$).

Since the events $A_1, A_2, \ldots A_m$ comprise a set, $\sum_{i=1}^{m} p_i = 1$.

* Sometimes instead of the number of sign changes there is examined the number of zeros, i.e., the intersection of the $\underline{i}(\underline{t})$ curve with the abscissa. It is clear that the two types of examination are equivalent to each other.

What is the probability that, as a result of n independent experiments, event A_1 occurs exactly k_1 times, event A_2 appears exactly k_2 times, etc., with event A_m appearing exactly k_m times? The answer to this question is provided by the formula

$$P_n(k_1, k_2, \ldots k_m) = \frac{n!}{k_1! k_2! \ldots k_m!} p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m}, \tag{1.55}$$

which is a generalization of the binomial formula (1.23). The numbers $k_1, k_2, \ldots k_m$ must, naturally, conform to the condition of

$$k_1 + k_2 + \ldots + k_m = n. \tag{1.56}$$

When $m = 2$ formula (1.55) turns into (1.23).

An example of a generalized system of independent tests can be found in the successive transmission of the letters of the Russian alphabet with the condition that each succeeding letter is transmitted independently of its predecessor. The result of a test is the appearance of one of 33 letters which, in the example at hand, constitute a set. The question can be asked, what is the probability that in a text of n letters the letter A occurs k_1 times, the letter B occurs k_2 times, etc., with the letter Ya occurring k_{33} times. The answer to this question can be obtained with the aid of formula (1.55), if there is available a table of probabilities (relative frequencies) p_i of the occurrence of the letters in the Russian language. A table of this sort was compiled by P. V. Prakhov, according to whose data, for instance, the probability of occurrence of the letter A is equal to $p_1 = 0.075$, the letter B -p_2 = 0.017, etc.

The example at hand also well illustrates the possibilities of the further generalization of a system of independent tests. The necessity of such a generalization is linked with the fact that in any language the sequence of letters is not independent.

In the simplest case it can be assumed that the occurrence of one of the 33 letters of the alphabet in a given test depends only on what letter occurred in the preceding test, and does not depend on what letters had occurred in earlier tests.

A succession of tests like that, for which the relative probability, in the (S + 1)th test, of the occurrence of one event of a set of events depends only on what event had occurred in the S-th test, and does not change as a result of additional information as to what events had occurred in earlier tests, is called a simple Markov chain. For homogenous Markov chains the relative probability P_{ij} of a transition from event A_i to event A_j depends only on these events, but does not depend on the number of the experiment. If the Russian language is to be regarded as a simple homogenous Markov chain, its statistical structure may be described as a collection of 33 a priori probabilities P_i of the occurrence of the i-th letter of the alphabet and of 33^2 transitions P_{ij}, i.e., a posteriori probabilities that the i-th letter of the alphabet will be followed by the j-th letter.

The next increase in complexity consists in the consideration of the probabilities of combinations of three letters, but not of more than three. The occurrence of a letter with a given test depends only on what letters were transmitted in the preceding two tests, but does not depend on what had occurred earlier. Here the relative probability must be P_{ijk}, i.e., the a posteriori probability that after the two-letter combination of the i-th and j-th letters of the alphabet there will follow the k-th letter of the alphabet. Continuing in such a manner, it is possible to obtain ever increasingly complex Markov chains.

Literature

1. B. V. Gnedenko. Kurs teorii veroyatnostey (Course in Probability Theory). Moskva, Gostekhizdat, 1954
2. G. P. Boyev. Teoriya veroyatnostey (Probability Theory). Moskva, Gostekhizdat, 1950
3. V. Feller. Vvedeniye v teoriyu veroyatnostey i yeye priloženiya (Introduction to Probability Theory and its Applications). Moskva, Foreign Literature Publishing House, 1952.
4. Ye.Ye. Slutskiy. Tablitsy dlya vychisleniya nepolnoy Γ- funktsii i funktsii veroy-

atnosti χ^2 (Tables for the computation of the incomplete Γ- function and the probability function of χ^2).

Moskva, pub. by AN SSSR (Academy of Sciences of the USSR) 1950.

5. V. I. Siforov. O metodakh rascheta nadezhnosti raboty sistem, soderzhashchikh bol'shoye chislo elementov (On the Methods of Calculating the Operational Reliability of Systems, Containing a Large Number of Components). "Izvestiya AN SSSR ("News of the Academy of Sciences of the USSR"), Department of Technical Sciences, 1954, No. 6.

6. A. N. Kolmogorov. Teoriya veroyatnostey (Matematika, yeye soderzhaniye, metody, i znacheniye. Tom II, gl. XI.) (Probability Theory (Mathematics, Its Content, Methods and Significance), Vol. II, Ch. XI). Pub. by AN SSSR 1956.

7. F. Vudvord (Woodward?), I. Devis (Davis?). Printsipy "obratnoy veroyatnosti" v teorii peredachi signalov (Principles of "Inverse Probability" in the Theory of Signal Transmission)(In the collection "Teoriya peredachi elektricheskikh signalov pri nalichii pomekh ("Theory of the Transmission of Electrical Signals with the Presence of Interference"), Moskva, For. Lit. Pub. Hse, 1953).

Chapter II

LAWS OF PROBABILITY DISTRIBUTION OF RANDOM VARIABLES

1. Definitions and Terminology

The subjects of investigation in the preceding chapter were random events, which either occurred or did not occur as the result of a random experiment. In practice it is frequently more convenient to designate the result of a random experiment by a certain real variable ξ , which is called a random variable. Thus, for instance, a result of the observation of noise in radio receivers in a fixed moment of time is characterized by a random momentary noise voltage. Another example of a random variable is the number of k occurrences of an event with n independent tests. As in the case of a single random experiment, it is impossible to predict the exact value of a random variable. It is possible only to predict the statistical rules of the change of a random variable, i.e., to assign probabilities to its values.

The range of possible values of a random variable may be finite or infinite, but countable (i.e., these values can be determined by means of a natural series of numbers). Then the random variable is called discrete. The possible values of a random variable may occupy an entire continuous range (finite or infinite), so that it is impossible to number these values. Then the random variable is called continuous.

Let the number of events, occurring in result of a random experiment, be finite. In accordance with this we will assign a certain simple number to each event. Then all the possible values of a certain discrete random variable will correspond to a set. The probability that a discrete random variable will take one of the possible values will be simply equal to the probability of occurrence of a random event corresponding to that value.

If the range of possible values of a continuous random variable is broken down into a finite number of segments, the aggregate of events wherein this variable falls

in any of these segments constitutes a set. Then the introduction of the concept of probability that the values of a random variable lie within the limits of a certain segment, is completely analogous to the discrete case. It obviously makes no sense to speak of the probability that a continuous random variable will take a given value, since this probability equals zero. In speaking of probability in connection with a continuous random variable, it is always necessary to bear in mind a certain segment of the complete range of its possible values.

In order to have a sufficiently complete understanding of the statistical rules characteristic of the random experiment at hand, it is thus necessary, first, to know the range of possible values of the random variable characterizing a random experiment, and second, the probabilities of these values.

The law that for each possible value of a discrete random variable or for some range of values of a continuous random variable, there is a corresponding probability that the random variable will take these values (or else will lie in some range of possible values), is called the law of the probability distribution of a random variable.

The analytical expressions of distribution laws are distribution functions, which for discrete random variables are functions of an integer argument, and for continuous random quantities are functions of a continuous argument.

Let us examine, for instance, the number of $\underline{k}$ occurrences of a random event with $\underline{n}$ independent tests. This number is a discrete random variable, the possible values of which are equal to $\underline{k} = 0,1,\ldots,\underline{n}$. As has been shown in the preceding chapter, the probability $\underline{P}_n(\underline{k})$ of the occurrence of an event $\underline{k}$ times with $\underline{n}$ independent tests equals

$$P\,(k)=\binom{n}{k}p^kq^{n-k}.$$

This formula establishes the law that for each possible value of $\underline{k}$ there is a corresponding probability $\underline{P}_n(\underline{k})$, i.e., the law of the probability distribution of the number of occurrences of an event with $\underline{n}$ independent tests. This law is frequently

called the binomial law of probability distribution.

2. Distribution Functions of Random Variables

Let us assume that the random variable ξ may take values within a range of from $-\infty$ to $+\infty$ and let us examine some level x, which can vary within the same limits. The present assumption does not restrict the generality of the results given below, since the variation of a random variable within a limited interval of values will simply signify that the probability of its reaching any zone of the numerical axis outside the indicated interval is equal to zero.

The function

$$F(x) = P(\xi \leqslant x). \tag{2.1}$$

which shows how the probability that the values of a random variable do not exceed the selected level of x depends on the magnitude of that level, is called the integral function of probability distribution.

Thus, for instance, for a discrete random variable which can with equal probability take any of the values of a natural series of numbers from 0 to n, the integral function of distribution will have the form (Fig. 3)

$$F(x) = P(\xi \leqslant x) = \begin{cases} 0, & x < 0 \\ \dfrac{x+1}{n+1}, & x = 0, 1, 2, \ldots n \\ 1, & x > n. \end{cases}$$

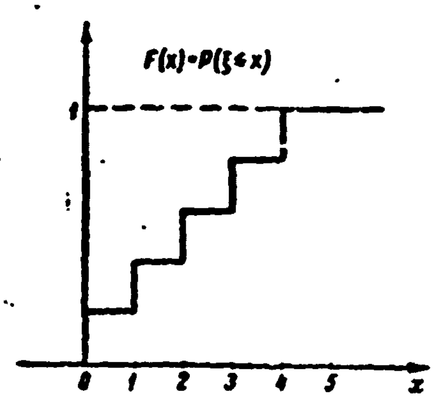


Fig. 3. Law of uniform distribution of a discrete random variable.

If the integral distribution function for a continuous random variable has the form (Fig. 4)

$$F(x)=P(\xi\leqslant x)=\begin{cases}0, & x<0\\ \frac{x}{h}, & 0\leqslant x\leqslant h\\ 1, & x>h,\end{cases}$$

it is said that this continuous random variable is evenly distributed over the interval (0, h). Examples of integral functions of distribution for a discrete random variable can also be found in formulas (1.42) and (1.53) of the preceding chapter, by means of which there is determined the probability that, with a large number n of independent experiments, the number of occurrences of an event does not exceed $m \leqslant n$. Distributions of the discrete random variables given by formulas (1.32) and (1.49) are called respectively the Laplace and Poisson distributions.

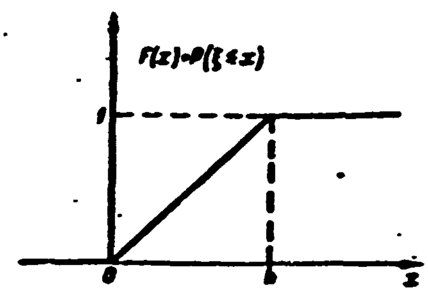


Fig. 4. Law of uniform distribution of a continuous variable.

Let us point out the basic properties of integral distribution functions. The values of these functions, which represent probabilities, must lie within the limits of from 0 to 1, where

$$F(-\infty)=P(\xi\leqslant-\infty)=0, \tag{2.2}$$

is the probability of an impossible event, and

$$F(\infty)=P(\xi\leqslant\infty)=1. \tag{2.3}$$

is the probability of a certain event.

The property expressed by equality (2.3) is analogous to the property of a set of events.

If the random variable lies below the level of $x_2 > x_1$, there exist two mutually exclusive possibilities: Either this random variable is located below level x_1, or it is located between levels x_1 and x_2. Then, using the rule of addition, we obtain

$$P(\xi < x_2) = P(\xi < x_1 \text{ or } x_1 < \xi < x_2) = P(\xi < x_1) + P(x_1 < \xi < x_2),$$

wherefrom

$$P(x_1 < \xi < x_2) = P(\xi < x_2) - P(\xi < x_1) \tag{2.4}$$

or, in view of (2.1),

$$P(x_1 < \xi < x_2) = F(x_2) - F(x_1). \tag{2.5}$$

Thus, the probability that a random variable is confined within definite limits is equal to the difference of the values of the integral distribution function for the upper and lower limits. Formula (1.38) in the preceding chapter is a special case of (2.5), when the Laplace function serves as the integral distribution function.

Since the left side of the equality (2.5) may not be negative,

$$F(x_2) > F(x_1) \text{ when } x_2 > x_1, \tag{2.6}$$

consequently, the integral function of distribution is always a non-diminishing function.

The integral distribution function of the discrete random variable ξ is discontinuous, it grows in steps with those values of x which are possible values of ξ. The integral distribution depicted in Fig. 3 is in this sense typical of the discrete random variable.

The integral distribution function of a continuous random variable is continuous and differentiable. The derivative

$$w(x) = \frac{dF(x)}{dx} \tag{2.7}$$

is called the probability density or the differential distribution function of a continuous random variable*. The probability density, as the derivative of a non-diminishing function (the integral distribution function) cannot take negative values, i.e.,

$$w(x) \geq 0. \tag{2.8}$$

The term "probability density" becomes understandable by examining the change in a continuous random variable within sufficiently narrow limits from x to $x + \Delta x$. Then

$$w(x) = \frac{dF(x)}{dx} \approx \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

or, considering (2.5), we find

$$w(x) \approx \frac{P(x < \xi \leq x + \Delta x)}{\Delta x}. \tag{2.9}$$

From expression (2.9) with a small Δx, it can be seen that $\omega(x)$ fulfills the meaning of probability density, since it is equal to the ratio of the probability of a random variable falling within the interval Δx, to the length of that interval.

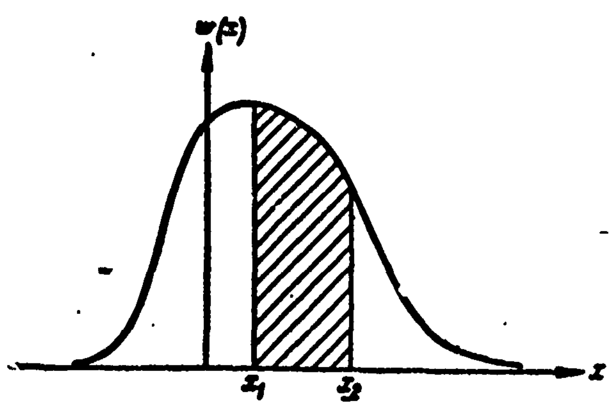


Fig. 5. Distribution curve of a continuous random variable. Shaded area is equal to the probability that values of the continuous random variable lie within the interval (x_1, x_2).

* From here on $\omega(x)$ is called simply the distribution function.

Let us represent the distribution function (probability density) in the form of a distribution curve (Fig. 5). Integrating both parts of (2.7) within the limits of from $-\infty$ to x, and bearing in mind (2.2), we obtain

$$P(\xi < x) = F(x) = \int_{-\infty}^{x} w(x)\,dx \tag{2.10}$$

and, bearing in mind (2.5), we obtain

$$P(x_1 < \xi < x_2) = \int_{x_1}^{x_2} w(x)\,dx. \tag{2.11}$$

Thus the probability that the values of a continuous random variable lie within a given interval is equal to the area under the distribution curve of that random variable for the interval in question (cf. shaded part of Fig. 5). In the special case with $x_1 = -\infty$ and $x_2 = \infty$, it follows from (2.11) that

$$\int_{-\infty}^{\infty} w(x)\,dx = 1, \tag{2.12}$$

i.e., the total area under the distribution curve is always equal to unity. In order to ensure the convergence of the improper integral (2.12), it is necessary that

$$\lim_{x\to-\infty} w(x) = \lim_{x\to\infty} w(x) = 0, \tag{2.13}$$

i.e., it is necessary that, in those cases when the random variable can vary within infinite limits, the abscissa in both directions be an asymptote of the distribution curve.

A distribution curve $\omega(x)$ may have one or more maxima. A maximum of $\omega(x)$ corresponds to a range of the most probable values of a random variable. The value of $\xi = x_m$, for which the probability density reaches a maximum, is called a mode.

It is simplest of all to check the above-enumerated properties of distribution functions (integral and probability density) on the example of uniform distribution (Fig. 4). The probability density in this case is, in accordance with (2.7) equal to

$$w(x)=\begin{cases}0, & x<0,\ x>h,\\ \frac{1}{h}, & 0<x<h,\end{cases}$$

i.e., maintains a constant value in the interval (0, h), thus justifying the appelation of uniform distribution.

One remark should be made concerning the designations employed. Random variables and the arguments of their functions are frequently designated by the same symbols, which can lead to misunderstandings. Therefore from here on random variables will be designated by Greek letters ξ, η, ζ ..., and the arguments of their distribution function by Latin letters, x, y, z.... Most frequently encountered in application is the so-called normal law of distribution of a random variable, to which the following section is devoted.

3. Normal Law of Distribution

The probability density of a continuous random variable which has a normal law of distribution depends on two parameters a and σ, and is determined by the formula

$$w(x)=c\,e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

The constant c cannot be arbitrary, but must be determined in such a manner as to satisfy condition (2.12), i.e.,

$$c\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx=1.$$

Since

$$\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \sigma \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sigma\sqrt{2\pi}^*,$$

it follows that

$$c = \frac{1}{\sqrt{2\pi}\sigma}$$

and, consequently,

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}. \tag{2.14}$$

Function (2.14) with $a = 0$ was used above with the asymptotic formula of Laplace [cf. formula (1.32)]. A table of the values of this function with $a = 0$ and $\sigma = 1$ is presented in Appendix II.

Fig. 6 shows the probability distribution curve for the normal law together with some parameters characteristic of it.** The curve has one maximum at the point $x = a$, whereby

$$w(a) = w_{max} = \frac{1}{\sqrt{2\pi}\sigma}. \tag{2.15}$$

At a distance of 3σ from the mode $x = a$ the probability density is only 0.004, and the area under the curve $\omega(x)$ in the band of $a \pm 3\sigma$ comprises 99.7% of the total area under the distribution curve, i.e., it may be considered with a probability of 0.997 that any random variable, distributed according to the normal law, lies between $a - 3\sigma$ and $a + 3\sigma$. (In the asymptotic approximation of Laplace this property was denoted by formula 1.45).

* Actually

$$\left\{\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz\right\}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx\,dy = \int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2}} r\,dr\,d\theta =$$

$$= 2\pi\int_0^{\infty} r e^{-\frac{r^2}{2}} dr = 2\pi e^{-\frac{r^2}{2}}\Big|_{\infty}^{0} = 2\pi.$$

** The curve $y = ce^{-\beta x^2}$ is called the Gaussian curve. The normal law of distribution is also often called the Gaussian law.

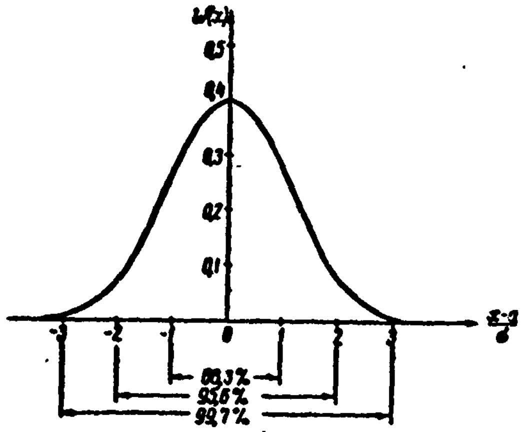


Fig. 6. Probability density curve for normal distribution (figures between arrows indicate the percentage of the total area under the distribution curve comprised by the area of the corresponding strip).

It is not difficult to show that the points of inflection, at which the distribution curve has a maximum slope, are determined from the equality

$$x = a \pm \sigma. \tag{2.16}$$

It can be seen from (2.15) and (2.16) that the larger the parameter σ, the smaller is the maximum probability density and the further is the point of inflection from the mode, i.e., the more widely dispersed are the significant values of function $\omega(\underline{x})$. Conversely, the smaller the parameter σ, the higher the maximum probability density, and the closer is the inflection point to the mode, i.e., the more compact are the significant values of the function $\omega(\underline{x})$, or the greater is the probability that a random variable, distributed according to the normal law, falls on a strip of a given width which includes the point $\underline{x} = \underline{a}$. Fig. 7 presents for comparison the probability density curves of the normal law of distribution for three values of σ. The envelope of this set of curves is the pair of hyperbolas

$$y = \pm \frac{1}{\sqrt{2\pi e}} \cdot \frac{1}{x - a},$$

whose asymptotes are the straight line x = a and the abscissa. When $\sigma \to \infty$ the distribution curve merges with the abscissa, and when $\sigma \to 0$ it turns into the delta-function*

$$\lim_{\sigma \to 0} w(x) = \lim_{\sigma \to 0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} = \delta(x-a).$$

(2.17)

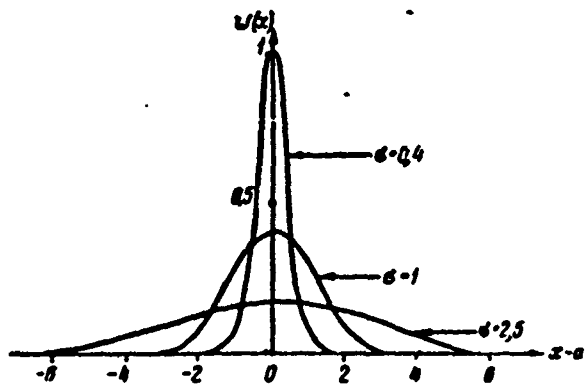


Fig. 7. Comparison of normally distributed probability density curves with various parameters σ.

The integral distribution function, i.e., the probability that a random variable, distributed according to the normal law lies below the level of x, is in accordance with (2.10) equal to

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = F\left(\frac{x-a}{\sigma}\right) \quad (2.18)$$

The function (2.18) with a = 0 and $\sigma = 1$ was examined in the preceding chapter in connection with the asymtpotic approximation of Laplace (1.37) for the binomial

* In view of the fact that the delta-function will in the future receive widespread use, some of the principal properties of this function are presented in Appendix IV.

law of distribution. (A table of the values of F(x) is shown in Appendix II.)

With the above-mentioned asymptotic approximation in mind, the normal law of distribution may be regarded as a limiting law, toward which tend the probability distributions of the numbers of occurrences of events, when the total number of independent tests grows without limit.

Fig. 8 shows a graph of the integral distribution function $F(x)$, which corresponds to function (2.18) with $a = 0$ and $\sigma = 1$. This function grows monotonously from 0 to 1, where, at the origin of the coordinates

$$F(0)=\frac{1}{2}F(\infty)=\frac{1}{2}.$$

The sum of the values of this function at two points, located symmetrically with respect to the origin of the coordinates (Figure 8) is equal to unity, which corresponds to the previously presented formula (1.39). Close to the origin, the integral distribution function $F(x)$ has a sector which comes close to being linear; it is well described by several initial terms of the exponential series

$$F(x)=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\left(x-\frac{x^3}{6}+\frac{x^5}{40}-\frac{x^7}{336}+\ldots\right). \tag{2.19}$$

With a sufficiently large argument there takes place the asymptotic expansion

$$F(x)\sim 1-\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\left(\frac{1}{x}-\frac{1}{x^3}+\frac{3}{x^5}-\frac{15}{x^7}+\ldots\right). \tag{2.20}$$

Let us note that if, in an alternating-signs series, we are to restrict ourselves to several terms, the error will be less than the value of the first rejected term. It follows from (2.20) that

$$1-\frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}<F(x)<1-\frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}}\left(1-\frac{1}{x^2}\right).$$

In Fig. 9 there are presented for comparison curves of the integral distribution function for the same values of σ as in Fig. 7. When $\sigma \to 0$, the integral distribution curves asymptotically approach the abscissa with $x < a$, and a parallel

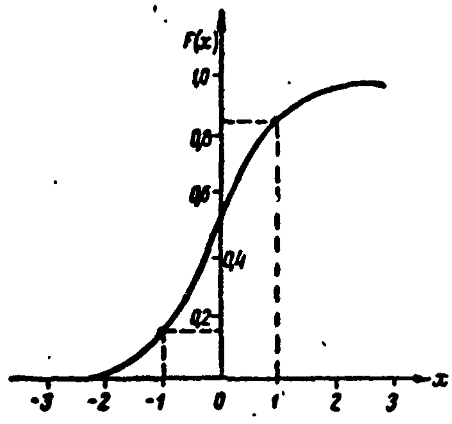


Fig. 8. Curve of the integral normal law of distribution.

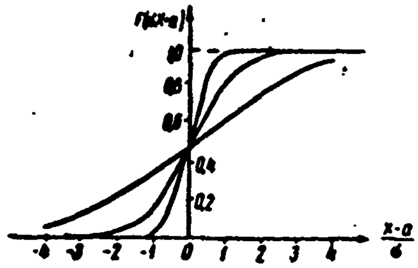


Fig. 9. Comparison of integral normal distribution curves with various parameters σ.

straight line, $F \equiv 1$, when $x > a$. The limiting curve has the form of a unit jump at the point $x = a$. This signifies that no matter how small is the magnitude of $\varepsilon > 0$, the probability that the random variable $\xi \leq a - \varepsilon$, is equal to zero, and it is certain that $\xi \geqslant a + \varepsilon$. Consequently, this random variable is the constant a.

The corresponding probability density of the constant a is equal to

$$w(x) = \delta(x - a).$$

It should be noted that the delta-function satisfies all the requirements laid down for distribution functions, in particular (cf. Appendix IV)

$$\int_{-\infty}^{\infty}\delta(x-a)\,dx=1.$$

Thus, constants may be regarded as random variables, whose distribution functions are delta-functions.

4. Joint Probability Distribution of Random Variables

Up to now we have discussed the probability distribution of only one random variable ξ . One may in an analogous manner examine the joint probability distribution of two dependent random variables

$$P(\xi\leqslant x,\ \eta\leqslant y)=F_2(x,\ y), \tag{2.21}$$

which yields the probability of combining the following two events: the random variable ξ is located below the level of x and the random variable η is located below the level of y. The function $F_2(x,y)$ is called the two-dimensional integral probability distribution function.

If in (2.21) it is assumed that $y = \infty$, then $F_2(x,\infty)$ will be equal to the probability that ξ is located below the level of x with any position of η , i.e., coincides with the integral distribution function of the random variable ξ

$$F_2(x,\ \infty)=F_{11}(x)^{*}, \tag{2.22}$$

Analogously,

$$F_2(\infty,\ y)=F_{12}(y). \tag{2.22'}$$

From (2.21) it follows also that

$$F_2(-\infty,\ y)=F_2(x,\ -\infty)=0 \tag{2.23}$$

---

* In such cases where a one-dimensional distribution function has a double index, the first figure indicates the order of the function, and the second figure, the number of the random variable whose distribution is determined by the function at hand.

and

$$F_2(\infty, \infty) = 1. \qquad (2.24)$$

It is possible to resort to a geometrical interpretation, regarding the variables ξ and η as coordinates of a point on a plane (Fig. 10). Then the two-dimensional integral distribution function $F_2(x,y)$ yields the value of the probability that a point with the random coordinates (ξ, η) will lie in that part of the plane which is shaded in Fig. 10. The probability that this point will lie within a rectangle with vertices at the points (x_1, y_1), (x_2, y_1), (x_2, y_2), (x_1, y_2), will equal

$$\begin{aligned} P(x_1 < \xi \leq x_2,\ y_1 < \eta \leq y_2) = \\ = F_2(x_1, y_1) - F_2(x_2, y_1) + \\ + F_2(x_2, y_2) - F_2(x_1, y_2). \end{aligned} \qquad (2.25)$$

Formula (2.25) generalizes formula (2.5) for a one-dimensional case; when $y_1 = -\infty$, $y_2 = \infty$ (2.25) turns into (2.5).

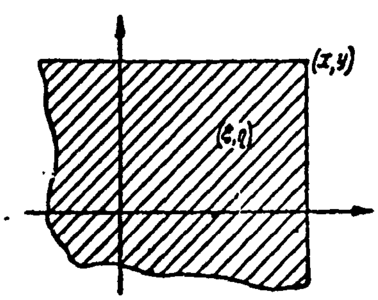


Fig. 10. Pair of random variables as coordinates of a point on a plane.

For continuous random variables, the two-dimensional integral function $F_2(x,y)$ is continuous. The mixed second-order derivative

$$w_2(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \qquad (2.26)$$

is called the two-dimensional probability density or the two-dimensional distribution function for the random variables ξ and η.

The two-dimensional distribution function is, just like the one-dimensional, always non-negative

$$w_2(x,\ y) \geqslant 0. \tag{2.27}$$

The two-dimensional distribution function is expressed in terms of the probability density by means of the double integral

$$F_2(x,\ y) = \int_{-\infty}^{x}\int_{-\infty}^{y} w_2(x,\ y)\,dx\,dy. \tag{2.28}$$

Let us represent the two-dimensional distribution function $\omega_2(x,y)$ as a surface (Fig. 11). Then the probability $P(A)$ that point A, with random coordinates whose probability density equals $\omega_2(x,y)$, falls within a region G of the plane, is equal to the volume bounded by the sector of surface $\omega_2(x,y)$, whose projection on the plane coincides with area G, i.e.,

$$P(A) = \iint_G w_2(x,\ y)\,dx\,dy. \tag{2.29}$$

The total volume under the distribution surface is, in accordance with (2.24) always equal to unity

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_2(x,\ y)\,dx\,dy = 1. \tag{2.30}$$

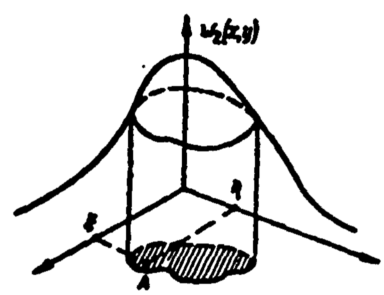


Fig. 11. Probability distribution surface of a continuous random variable.

In order to ensure the convergence of the improper integral of (2.30), it is necessary that the surface $\omega_2(x,y)$ asymptotically approach the plane of (x,y) in all directions.

By means of a given two-dimensional density $\omega_2(x,y)$ it is not difficult to find the probability distribution laws of each of the random variables. In fact, from (2.22) and (2.28) we find for the random variable ξ

$$F_{11}(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty}w_2(x,\ y)\,dy\,dx,=\int_{-\infty}^{x}w_{11}(x)\,dx, \tag{2.31}$$

where

$$w_{11}(x)=\int_{-\infty}^{\infty}w_2(x,\ y)\,dy. \tag{2.32}$$

Analogously for the random variable η

$$F_{12}(y)=\int_{-\infty}^{y}\int_{-\infty}^{\infty}w_2(x,\ y)\,dx\,dy=\int_{-\infty}^{y}w_{12}(y)\,dy, \tag{2.33}$$

where

$$w_{12}(y)=\int_{-\infty}^{\infty}w_2(x,\ y)\,dx. \tag{2.34}$$

Let us present separately the formulas, important to practical applications, of the joint distribution functions of two independent random variables.

The random variables ξ and η are called independent, if

$$F_2(x,\ y)=P(\xi<x,\ \eta<y)=F_{11}(x)F_{12}(y), \tag{2.35}$$

where $F_{11}(x)$ and $F_{12}(y)$ are integral distribution functions of each of the random variables under examination. An analogous formula holds true also for the two-dimensional probability density

$$w_2(x,\ y)=w_{11}(x)\cdot w_{12}(y), \tag{2.36}$$

where $\omega_2(x)$ and $\omega_{12}(y)$ are distribution functions of each of the random variables under examination.

Thus, two random variables are independent, if their joint distribution function is the product of the individual distribution functions.

The simplest example of a two-dimensional distribution function

$$w_2(x, y)=\begin{cases} \frac{1}{(b-a)(d-c)}, & a<x<b, \quad c<y<d, \\ 0 & , x<a, \ x>b, \ y<c, \ y>d \end{cases}$$

corresponds to the uniform distribution of two random variables in the rectangle $G(a,b,c,d)$. The probability that point A with the coordinates ξ and η falls into a region g, lying within rectangle G, will be equal in this case simply to the ratio of the areas of region g and rectangle G.

The study of the joint distribution laws of random variables may be generalized for n random variables $\xi_1, \xi_2, \ldots \xi_n$. In this case the n-dimensional integral distribution function $F_n(x_1, x_2, \ldots x_n)$ yields the probability that the random variable ξ_1, x_1 and also that the random variable ξ_2 lies below the level of x_2, and the random variable ξ_n lies below the level of x_n, etc., i.e.,

$$P(\xi_1<x_1, \xi_2<x_2, \ldots \xi_n<x_n)=F_n(x_1, x_2, \ldots x_n). \tag{2.37}$$

The random variables $\xi_1, \xi_2, \ldots \xi_n$ are called mutually independent if

$$F_n(x_1, x_2, \ldots x_n)=F_{11}(x_1)F_{12}(x_2)\ldots F_{1n}(x_n), \tag{2.38}$$

where $F_{1k}(x_k)$ is the integral distribution function of the random variable ξ_k ($k = 1,2,\ldots n$).

Formula (2.38) generalizes the previously presented formula (2.35).

If m of n arguments of function (2.37) turn to infinity, i.e., if it is known that m of n random values may be arbitrary, then this function becomes the integral distribution function of the remaining $n - m$ random variables

$$F_n(\infty, \ldots \infty, x_{m+1}, \ldots x_n)=F_{n-m}(x_{m+1}, \ldots x_n). \tag{2.39}$$

From the definition (2.37) it follows that

$$F_n(-\infty, x_2, \ldots x_n) = F_n(x_1, -\infty, \ldots x_n) = \ldots = \\ = F_n(x_1, x_2, \ldots -\infty) = 0 \tag{2.40}$$

and

$$F_n(\infty, \infty, \ldots \infty) = 1. \tag{2.40'}$$

If the n-dimensional integral function has the derivative

$$w_n(x_1, x_2, \ldots x_n) = \frac{\partial^n F_n(x, x_2, \ldots x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}, \tag{2.41}$$

then this derivative is called the n-dimensional density distribution or the n-dimensional distribution function of the random variables $\xi_1, \xi_2, \ldots \xi_n$.

The integral distribution function (2.37) is expressed in terms of the probability density by means of the n-multiple integral

$$F_n(x_1, x_2, \ldots x_n) = \\ = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\ldots\int_{-\infty}^{x_n} w_n(x_1, x_2, \ldots x_n)\, dx_1 dx_2 \ldots dx_n. \tag{2.42}$$

The aggregate of the random values $\xi_1, \xi_2, \ldots \xi_n$ may be treated as the coordinates of a point or as the components of a vector in n-dimensional space. The probability P(A) that point A with random coordinates, whose probability density is equal to $\omega_n(x_1, x_2, \ldots x_n)$, falls in the region G of the space, equals

$$P(A) = \iint_G \ldots \int w_n(x_1, x_2, \ldots x_n)\, dx_1 dx_2 \ldots dx_n. \tag{2.43}$$

Making use of (2.40'), we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} w_n(x_1, x_2, \ldots x_n)\, dx_1 dx_2 \ldots dx_n = 1. \tag{2.44}$$

If the random variables $\xi_1, \xi_2, \ldots \xi_n$ are mutually independent, it follows from (2.38) that

$$w_n(x_1, x_2, \ldots x_n) = w_{11}(x_1) w_{12}(x_2) \ldots w_{1n}(x_n). \tag{2.45}$$

where $\omega_{1k}(x_k)$ is the probability density of the random value ξ_k $(k = 1,2,\ldots n)$.

It follows from formulas (2.39) and (2.42) that, with $m \leqslant n$

$$F_n(x_1, x_2, \ldots x_m, \infty, \ldots \infty) = \\ = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_m} w_m(x_1, x_2, \ldots x_m)\, dx_1\, dx_2 \ldots dx_m, \tag{2.46}$$

where

$$w_m(x_1, x_2, \ldots x_m) = \\ = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-m \text{ раз}} w_n(x_1, x_2, \ldots x_n)\, dx_{m+1} \ldots dx_n. \tag{2.47}$$

Thus, knowing the n-dimensional distribution function, it is always possible to obtain the distribution function of any group of m random variables $(1 \leqslant m \leqslant n)$ by means of integrating within the infinite limits the n-dimensional function with respect to the remaining n-m variables.

5. Multidimensional Normal Law of Distribution

The n-dimensional normal law of distribution provides an example of the joint probability distribution of a system of random variables $\xi_1, \xi_2, \ldots \xi_n$ which is of the greatest importance in practical application. The n-dimensional distribution function corresponding to the normal law is by definition equal to

$$w_n(x_1, x_2, \ldots x_n) = \frac{1}{\sigma_1 \sigma_2 \ldots \sigma_n \sqrt{(2\pi)^n D}}\, e^{-\frac{1}{2D} \sum_{i=1}^{n} \sum_{k=1}^{n} D_{ik} \frac{x_i - a_i}{\sigma_i} \frac{x_k - a_k}{\sigma_k}}, \tag{2.48}$$

where D is a determinant of the n-th order which has the following form

$$D = \begin{vmatrix} 1 & r_{12} & \ldots & r_{1n} \\ r_{21} & 1 & \ldots & r_{2n} \\ \cdot & \cdot & \cdot\,\cdot\,\cdot & \vdots \\ r_{n1} & r_{n2} & \ldots & 1 \end{vmatrix}, \quad r_{ik} = r_{ki}, \quad |r_{ik}| \leqslant 1, \tag{2.49}$$

and the quantity D_{ik} is the co-factor of the element r_{ik} in the determinant D.

Thus the n-dimensional distribution function (2.48) depends on the 2 n parameters a_i, σ_i and on the $\frac{n(n-1)}{2}$ parameters r_{ik}, i.e., on a total of $\frac{n^2 + 3n}{2}$ parameters.

If the random variables $\xi_1, \xi_2, \ldots \xi_n$ are independent and each of them is distributed according to the normal law with the respective parameters a_k, σ_k (k = 1, 2,...n), it follows from (2.14) and (2.45) that the joint n-dimensional distribution of these random variables is determined by the function

$$w_n(x_1, x_2, \ldots x_n) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-a_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-a_2)^2}{2\sigma_2^2}} \cdots$$

$$\cdots \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(x_n-a_n)^2}{2\sigma_n^2}},$$

or

$$w_n(x_1, x_2, \ldots x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1\sigma_2 \ldots \sigma_n} e^{-\frac{1}{2}\sum_{k=1}^{n} \frac{(x_k-a_k)^2}{\sigma_k^2}}. \qquad (2.50)$$

Formula (2.50) is a special case of the general formula (2.48) with $r_{ik} = 0$ ($i \neq k$), for which $D = 1$, $D_{ik} = 0$ when $i \neq k$, and $D_{ik} = 1$ with $i = k$. When $n = 1$ formula (2.50) turns into (2.14).

Let us now examine in greater detail the normal law of distribution of two random variables. A two-dimensional distribution function depends in this case on 5 parameters: $a_1, \sigma_1, a_2, \sigma_2, r_{12} = r$. The determinant D is equal to

$$D = \begin{vmatrix} 1 & r \\ r & 1 \end{vmatrix} = 1 - r^2,$$

and the co-factors $D_{11} = D_{22} = 1$, $D_{12} = D_{21} = r$.

From the general formula (2.48) we obtain, when $n = 2$, the following expression for the two-dimensional distribution function of two normally distributed random values

$$w_2(x_1, x_2) =$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left\{\frac{(x_1-a_1)^2}{\sigma_1^2} - 2r\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right\}}. \qquad (2.51)$$

The surface of a two-dimensional normal distribution which corresponds to (2.51) has the form depicted in Fig. 12. At the point $x_1 = a_1$, $x_2 = a_2$ the probability density is at its maximum and is equal to

$$w_2(a_1, a_2) = w_{2max} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}. \tag{2.52}$$

From this point the distribution surface approaches the plane (x_1, x_2) monotonously from all directions. Any vertical intersection of the surface $z = \omega_2(x_1, x_2)$ with a plane passing through the point of maximum probability density represents a Gaussian curve (cf. note on p. 46). The probability density maintains constant values along the ellipses which are horizontal sections of surface (2.51). The equation of a family of ellipses of equal probability densities has the form of

$$\varphi(x_1, x_2) = \frac{(x_1 - a_1)^2}{\sigma_1^2} - 2r\frac{(x_1 - a_1)(x_2 - a_2)}{\sigma_1\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} = \lambda^2. \tag{2.53}$$

Various but constant, for a given λ , values of the probability density correspond to various values of the parameter $\lambda > 0$ for all points of ellipse (2.53). When $\lambda > 0$ the ellipse degenerates into the point $x_1 = a_1$, $x_2 = a_2$ of maximum density, and as λ increases the intersecting horizontal plane is lowered ever more, the density on the ellipses of equal probabilities undergoing a corresponding decrease. The lengths of the large and small half-axes of an ellipse of equal probabilities are proportional to the parameter λ , and are equal to

$$\beta = \frac{\lambda\sigma_2}{\sqrt{1-r}}, \quad \alpha = \frac{\lambda\sigma_1}{\sqrt{1+r}}. \tag{2.54}$$

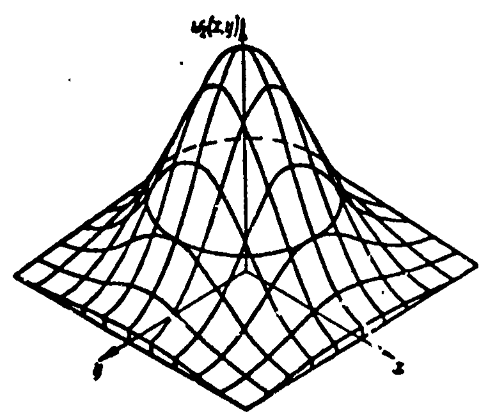


Fig. 12. Surface of two-dimensional normal law of distribution.

The probability that a point of the plane ($\underline{x}_1$, $\underline{x}_2$) with random coordinates, distributed according to the two-dimensional normal law, will lie within an ellipse of equal probabilities with a fixed parameter λ, is equal, according to (2.43), to

$$P(\lambda)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\iint\limits_{g(\lambda)} e^{-\frac{\varphi(x_1,x_2)}{2(1-r^2)}}\,dx_1dx_2, \tag{2.55}$$

where $\underline{g}$ (λ) is the area of the plane ($\underline{x}_1$, $\underline{x}_2$) bounded by the ellipse (2.53).

The integral (2.55) is computed by resorting to polar coordinates. As a result of such a computation we obtain

$$P(\lambda)=1-e^{-\frac{\lambda^2}{2(1-r^2)}}.$$

Expressing the parameter λ in terms of the lengths of the half-axes of the equal-probability ellipse, we find also

$$P(\lambda)=1-e^{-\frac{a^2}{2\sigma_1\sigma_2\sqrt{1-r^2}}}. \tag{2.56}$$

It is not difficult to show that each of the mutually dependent random variables, distributed according to the two-dimensional normal law, is also distributed normally.

Substituting (2.51) into (2.32), we find

$$w_{11}(x_1)=\int\limits_{-\infty}^{\infty} w_2(x_1,x_2)\,dx_2=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{(x_1-a_1)^2}{2(1-r^2)\sigma_1^2}}\times$$

$$\times\int\limits_{-\infty}^{\infty}\exp\left\{\frac{-1}{2(1-r^2)}\left[\frac{(x_2-a_2)^2}{\sigma_2^2}-2r\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2}\right]\right\}dx_2.$$

Completing the square in the expression in brackets, we obtain

$$w_{11}(x_1)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x_1-a_1)^2}{2\sigma_1^2}}\times$$

$$\times\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{x_2-a_2}{\sigma_2}-r\frac{x_1-a_1}{\sigma_1}\right)^2\right\}dx_2,$$

wherefrom

$$w_{11}(x_1)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x_1-a_1)^2}{2\sigma_1^2}}. \tag{2.57}$$

Analogously

$$w_{12}(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-a_2)^2}{2\sigma_2^2}}. \tag{2.58}$$

If the random variables are independent, then in accordance with (2.50)

$$w_2(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(x_1-a_1)^2}{\sigma_1^2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right]}. \tag{2.59}$$

Formula (2.59) is a special case of (2.51) when $r = 0$. When $\sigma_1 = \sigma_2 = \sigma$ the ellipses of equal probability turn into circles with radius $\rho = \lambda\sigma$.

The probability that a point on a plane, whose coordinates are independent, normally distributed random variables, lies within a circle of radius ρ, whose center is at point (a_1, a_2), is according to (2.56)

$$P(\rho) = 1 - e^{-\frac{\rho^2}{2\sigma^2}}, \ \rho > 0. \tag{2.60}$$

In this case it is, as can be seen from (2.60), practically certain that a point with the indicated coordinates will fall in a circle with a radius of 3σ whose center is at the point $x_1 = a_1$, $x_2 = a_2$.

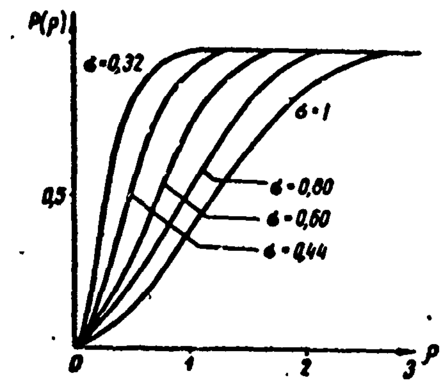


Fig. 13. Rayleigh's Integral Distribution Function

The probability density, which corresponds to the integral distribution function (2.60), equals

$$w(\rho) = \frac{dP(\rho)}{d\rho} = \frac{\rho}{\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}}, \ \rho > 0. \tag{2.61}$$

Distribution function (2.61), which is often called the Rayleigh function,

determines the probability that two independent random variables, distributed in accordance with the normal law having equal parameters $\sigma_1 = \sigma_2 = \sigma$, fall within a ring bounded by the two concentric circles of radii ρ and $\rho + d\rho$, whose center lies at the point of maximum joint probability density of these random variables.

Fig. 13 shows several curves of the integral Rayleigh distribution function (2.60), and Fig. 14 - the respective probability density curves (2.61). The maximum density $\omega_{max} = \frac{1}{\sigma\sqrt{e}}$ is attained when $\rho = \sigma$. This is the point of inflexion of the integral distribution curve. The slope of the probability density curve at the origin is equal to $\omega'(0) = \frac{1}{\sigma^2}$. The envelope of a set of Rayleigh distribution function curves is a hyperbola, whose equation is $y = \frac{2}{e\rho}$.

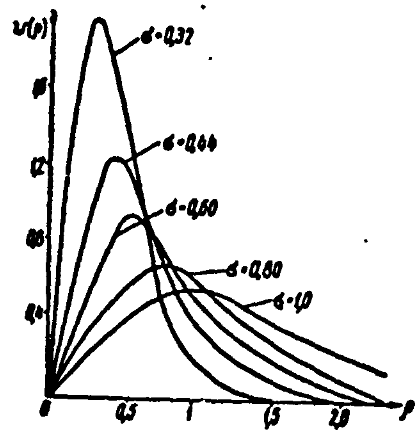


Fig. 14. Probability Density for a Rayleigh Distribution

Let us present in conclusion the expression of the joint probability distribution function of three normally distributed random variables. From the general formula (2.48) when $\underline{n} = 3$ we obtain

$$w_3(x_1, x_2, x_3) = \frac{e^{-M}}{\sigma_1\sigma_2\sigma_3(2\pi)^{3/2}\sqrt{1 - r_{12}^2 - r_{23}^2 - r_{31}^2 + 2r_{12}r_{23}r_{31}}}, \tag{2.62}$$

where

$$M = \frac{1}{2(1 - r_{12}^2 - r_{23}^2 - r_{31}^2 + 2r_{12}r_{23}r_{31})}\left[\frac{1 - r_{23}^2}{\sigma_1^2}(x_1 - a_1)^2 + \right.$$

$$+ \frac{1 - r_{31}^2}{\sigma_2^2}(x_2 - a_2)^2 + \frac{1 - r_{12}^2}{\sigma_3^2}(x_3 - a_3)^2 -$$

$$- \frac{2(r_{12} - r_{23}\cdot r_{31})}{\sigma_1\sigma_2}(x_1 - a_1)(x_2 - a_2) -$$

$$- \frac{2(r_{23} - r_{31} \cdot r_{12})}{\sigma_2 \sigma_3} (x_2 - a_2)(x_3 - a_3) - - \frac{2(r_{31} - r_{12} \cdot r_{23})}{\sigma_3 \sigma_1} (x_3 - a_3)(x_1 - a_1) \Big]. \quad (2.63)$$

A three-dimensional distribution function $\omega_3(x_1, x_2, x_3)$ depends on as many as 9 parameters: a_1, a_2, a_3, σ_1, σ_2, σ_3, r_{12}, r_{23}, r_{31}.

6. Numerical Characteristics of a Random Variable.

The integral distribution function (or its derivative, the probability density), examined in the preceding sections, provides a complete characterization of a random variable, as it points out the range in which the values of the random variable change, and the probability of finding it in specific sections of that range. However, in a number of cases it is sufficient to know much less about a random variable, an overall concept being adequate.

This is analogous to the case when instead of a description of the most minute details of a solid, one deals only with such of its overall numerical characteristics as length, width, height, volume, moment of inertia, etc..

In the probability theory, certain numerical constants, obtained according to specific rules from the distribution functions, serve as overall characteristics of a random variable.

As such numerical characteristics of random variables it is customary to consider the so-called distribution moments of various orders. For continuous random variables the distribution moments of the k-th order are determined according to the formula*

$$m_k\{\xi\} = \int_{-\infty}^{\infty} x^k w(x)\, dx \quad (2.64)$$

under the assumption that the improper integral converges. The numbers, m_k, may be treated geometrically as moments of inertia of the corresponding orders of a plane figure, bounded by the abscissa and the curve $y = \omega(x)$.

---

* The symbol $m_k(\xi)$ denotes not a function of the random variable ξ, but an averaging operation of the variable ξ^k from the totality of its possible values (compare p. 104, formula 3.48).

If the random variable is discrete and takes the values of $x_1, x_2, \ldots x_n$ with the probabilities of $p_1, p_2, \ldots p_n$, its k-th moment of distribution is equal to

$$m_k\{\xi\}=\sum_{r=1}^{n} x_r^k p_r. \tag{2.65}$$

The simplest numerical characteristic of a random variable - a distribution moment of the first order, determining the abscissa of the gravitational center of a distribution curve, is called the mathematical expectation or the mean value of a random variable. In accordance with the general definition, the mean value of a continuous random variable is equal to

$$m_1\{\xi\}=\int_{-\infty}^{\infty} xw(x)\,dx, \tag{2.66}$$

and the mean value of a discrete random variable is equal to

$$m_1\{\xi\}=\sum_{r=1}^{n} x_r p_r. \tag{2.67}$$

The difference

$$\Delta_\xi=\xi-m_1\{\xi\}$$

between a random variable and its mean value is called the deviation of a random variable. The deviation of a random variable is equal to the distance from the abscissa of the gravitational center of the distribution curve of this random variable. Distribution moments of the deviation probabilities of a random variable are called central moments and are designated by $M_k\{\xi\}$. In accordance with (2.64).

$$M_k\{\xi\}\doteq m_k\{\xi-m_1\}=\int_{-\infty}^{\infty}(x-m_1)^k w(x)\,dx. \tag{2.68}$$

In distinction from the moments $m_k\{\xi\}$ of a distribution curve with respect to the coordinate axis, which are called initial moments (or, simply, moments [tr.]), central moments are the moments of the distribution curve with respect to the axis passing through the gravitational center of this curve.

For a discrete random variable the central moments of distribution are defined respectively by the summation

$$M_k\{\xi\}=\sum_{r=1}^{n}(x_r-m_1)^k p_r. \qquad (2.69)$$

If the mean value of a random variable is equal to zero, then $\Delta_\xi = \xi$ and the central moments of distribution coincide with the moments. It is obvious that a first central moment is always equal to zero.

In accordance with the general definition, a first moment for a continuous random variable is equal to

$$m_2\{\xi\}=\int_{-\infty}^{\infty} x^2 w(x)\,dx \qquad (2.70)$$

and, for a discrete random variable, to

$$m_2\{\xi\}=\sum_{r=1}^{n} x_r^2 p_r. \qquad (2.71)$$

A second central moment of distribution is called the dispersion of the random variable ξ, and is defined in accordance with (2.68) and (2.69) by the formulas

$$M_2\{\xi\}=\int_{-\infty}^{\infty}(x-m_1)^2 w(x)\,dx \qquad (2.72)$$

for a continuous random variable and

$$M_2\{\xi\}=\sum_{r=1}^{n}(x_r-m_1)^2 p_r \qquad (2.73)$$

for a discrete random variable.

The magnitude $\sqrt{M_2}$ is called the mean quadratic or standard deviation of the random variable ξ from its mean value. In physical applications it is customary to call the square root of the dispersion, the fluctuation of a random variable. The ratio of the deviation of a random variable to the standard deviation is called the normalized deviation of a random variable.

The central second moment and the initial second moment are not independent. They are linked by a relationship which follows directly from (2.72) [and, analogously for a discrete random variable, from (2.73)]:

$$M_2 = \int_{-\infty}^{\infty} (x^2 - 2xm_1 + m_1^2)\, w(x)\, dx =$$
$$= \int_{-\infty}^{\infty} x^2 w(x)\, dx - 2m_1 \int_{-\infty}^{\infty} xw(x)\, dx + m_1^2 \int_{-\infty}^{\infty} w(x)\, dx,$$

and since

$$\int_{-\infty}^{\infty} xw(x)\, dx = m_1, \quad \int_{-\infty}^{\infty} w(x)\, dx = 1,$$

then

$$M_2 = m_2 - m_1^2. \tag{2.74}$$

Thus the dispersion of a random variable is equal to the difference between its second moment of distribution and the square of its mean value.

In the following sections there are examined properties of the mean value and of dispersion, and these numerical characteristics are defined for the distribution functions examined above.

7. Examples of Computation of Mean Value and Dispersion of a Random Variable.

Let us find the mean value and the dispersion of a discrete random variable, distributed according to the binomial law (1.23). Making use of formula (2.67), we find*

$$m_1 = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k},$$

and since

$$\binom{n}{k} = \binom{n-1}{k-1} \frac{n}{k},$$

then

$$m_1 = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)} =$$
$$= np \sum_{r=0}^{n-1} \binom{n-1}{r} p^r q^{n-1-r}.$$

As it represents the sum of the probabilities of a set of events, the latter summation is equal to 1 and, consequently,

$$m_1 = np. \tag{2.75}$$

*The summation starts at $k = 1$, since, when $k = 0$, the corresponding term in the summation is equal to zero.

Comparing (2.75) with (1.27), we conclude that the mean value m_1 of the numbers of occurrences of an event, with n independent tests, differs from the most probable number of k_0 occurrences by less than one.

The dispersion of a discrete, binomially distributed random variable is formed according to formula (2.74). For this we first determine the initial second moment

$$m_2 = \sum_{k=1}^{n} k^2 \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} q^{n-1-k+1} =$$
$$= np \sum_{r=0}^{n-1} (r+1) \binom{n-1}{r} p^r q^{n-r-1} =$$
$$= np \sum_{r=0}^{n-1} r \binom{n-1}{r} p^r q^{n-r-1} + np \sum_{r=0}^{n-1} \binom{n-1}{r} p^r q^{n-r-1}.$$

The first summation equals $(n-1)p$, as the average number of occurrences of an event with $(n - 1)$ tests, and the second equals the sum of the probabilties of a set of events. Therefore

$$m_2 = np(n-1)p + np.$$

Then the value of the dispersion M_2 of the occurrence number of events with n independent tests is equal to

$$M_2 = np(n-1)p + np - (np)^2 = np - np^2 = np(1-p)$$

or

$$M_2 = npq. \qquad (2.76)$$

Comparing (2.76) with (1.30') we find that the parameter σ^2 in the Laplace asymptotic formula is the dispersion of the number of occurrences of an event with n independent tests.

The mean value of a continuous random variable, distributed according to the normal law (2.14), is in accordance with (2.66) equal to

$$m_1 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (u+a) e^{-\frac{u^2}{2\sigma^2}} du.$$

The integral

$$\int_{-\infty}^{\infty} u e^{-\frac{u^2}{2\sigma^2}} du = 0,$$

since the integrand function is odd, and the integral

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\sigma^2}} du = 1$$

by virtue of (2.12). Therefore the mean value of a normally distributed random variable is equal to

$$m_1 = a. \tag{2.77}$$

Thus the mean value of a random variable, distributed according to the normal law, is equal to the parameter a of that law.

The dispersion of a normally distributed, continuous random variable is, in accordance with (2.72), equal to

$$M_2 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}} dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2\sigma^2}} du = \frac{-\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \, d e^{-\frac{u^2}{2\sigma^2}},$$

and, after integrating by parts, we obtain

$$M_2 = \sigma^2. \tag{2.78}$$

Thus the significance of both parameters of the normal distribution function has been determined; one of them is the mean value of a random variable, and the other is its dispersion.

The mean value and dispersion of a continuous, Rayleigh-distribution (2.61) random variable can be found with equal ease

$$m_1 = \frac{1}{\sigma^2} \int_0^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx = \sigma \sqrt{\frac{\pi}{2}}, \tag{2.79}$$

$$M_2 = \frac{1}{\sigma^2} \int_0^{\infty} x \left(x - \sigma \sqrt{\frac{\pi}{2}}\right)^2 e^{-\frac{x^2}{2\sigma^2}} dx = 2\sigma^2 - \frac{\pi\sigma^2}{2} = \frac{4-\pi}{2}\sigma^2. \tag{2.79'}$$

8. Coefficients of Asymmetry and Excess

From the definitions provided in Section 6 and also from the examples cited in Section 7 it follows that the mean value of a random variable provides a conception of the domain of its most probable values, and the dispersion points out the extent to which the values of the random variable are scattered with respect to its mean value. In other words, the mean value and the dispersion are the characteristics of location and scattering of the law of distribution of a random variable. However, these numerical characteristics do not reflect all the specific properties of a distribution curve.

One such property is the symmetry or asymmetry of a distribution curve with respect to the axis passing through its gravitational center. In any symmetrical distribution, an example of which is the normal law of distribution, every odd central moment is equal to zero; this follows directly from (2.68). Therefore the simplest of the odd moments, the third central moment, may in the first approximation serve as an asymmetry characteristic of the law of distribution.

The third central moment

$$M_3=\int_{-\infty}^{\infty}(x-m_1)^3 w(x)\,dx \tag{2.80}$$

may be expressed in terms of the first three moments similarly to how the dispersion is expressed in (2.74) in terms of the first and second moments.

Cubing the binomial in the integrand of (2.80), we find

$$M_3=m_3-3m_1m_2+2m_1^3. \tag{2.81}$$

In mathematical statistics, it is customary to characterize the asymmetry of a distribution curve by the dimensionless expression

$$k=\frac{M_3}{\sqrt{M_2^3}}, \tag{2.82}$$

which is called the coefficient of asymmetry.

An example of asymmetrical distribution is the Rayleigh distribution (2.61).

The third moment for this distribution is equal to

$$m_3=\frac{1}{\sigma^2}\int_0^\infty x^4 e^{-\frac{x^2}{2\sigma^2}}dx=3\sigma^3\sqrt{\frac{\pi}{2}}.$$

Therefore

$$M_3=3\sigma^3\sqrt{\frac{\pi}{2}}-3\sigma\sqrt{\frac{\pi}{2}}\cdot 2\sigma^2+2\sigma^3\left(\sqrt{\frac{\pi}{2}}\right)^3=$$
$$=(\pi-3)\sqrt{\frac{\pi}{2}}\sigma^3,$$

and, consequently, the coefficient of asymmetry of the Rayleigh law of distribution is equal, in accordance with (2.82), to

$$k=2\sqrt{\frac{\pi}{4-\pi}}\cdot\frac{\pi-3}{4-\pi}=0,63. \tag{2.83}$$

As a characteristic of the smoothness of a distribution curve about its mode, there is employed the dimensionless coefficient of excess

$$\gamma=\frac{M_4}{M_2^2}-3. \tag{2.84}$$

For a normal law of distribution

$$M_4=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-a)^4 e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}u^4 e^{-\frac{u^2}{2\sigma^2}}du=$$
$$=-\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u^3 de^{-\frac{u^2}{2\sigma^2}}=\frac{3\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}u^2 e^{-\frac{u^2}{2\sigma^2}}du=3\sigma^2 M_2,$$

and, since $M_2=\sigma^2$, according to (2.84)

$$\gamma=\frac{3\sigma^4}{\sigma^4}-3=0.$$

Thus the normal distribution curve has a zero coefficient of excess. A positive value of γ indicates that in the vicinity of the mode the distribution curve has a higher and sharper peak than a normal distribution curve with the same gravitational center and dispersion. A negative value of the coefficient of excess indicates a flatter peak than that of the corresponding normal distribution curve.

Analogously to M_2 and M_3, the fourth central moment is expressed in terms of the initial moments. This expression has the form of

$$\overset{\cdot\cdot}{M}_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4. \tag{2.85}$$

Utilization of (2.85) facilitates the computation of the coefficient of excess.

Let us, for instance, determine the coefficient of excess for the Rayleigh distribution (2.61). We find first the magnitude

$$m_4 = \frac{1}{\sigma^2}\int_0^\infty x^5 e^{-\frac{x^2}{2\sigma^2}}\,dx = 8\sigma^4$$

and, keeping in mind the values obtained for m_1, m_2, and m_3 of the Rayleigh distribution, we obtain by means of (2.85)

$$M_4 = 8\sigma^4 - 4\cdot 3\sigma^3\sqrt{\frac{\pi}{2}}\cdot\sigma\sqrt{\frac{\pi}{2}} + 6\cdot 2\sigma^2\cdot\sigma^2\frac{\pi}{2} - \\ - 3\sigma^4\cdot\frac{\pi^2}{4} = \left(8 - \frac{3\pi^2}{4}\right)\sigma^4.$$

The coefficient of excess is, according to (2.48), equal to

$$\gamma = \frac{32 - 3\pi^2}{(4-\pi)^2} - 3 \approx -0,3,$$

which indicates the flatter and lower character of the Rayleigh distribution curve in comparison to a normal distribution curve with a mean value of $\sigma\sqrt{\frac{\pi}{2}}$ and a dispersion of $2\sigma^2$.

Let us note that if the form of the distribution function is given, but the values of the parameters are unknown, these unknown parameters are determined by means of the moments of distribution. Thus, for instance, both parameters of the normal law of distribution are determined by the mean value and dispersion. If the distribution function depends on a large number of parameters, for their determination there may be necessary a knowledge of the value of the coefficients of asymmetry and excess, or even of moments of a higher order than the fourth. However, if the form of the distribution function is unknown, the knowledge of some number of distribution moments does not, generally speaking, provide a possibility of precisely determining the unknown distribution function.

9. Numerical Characteristics of Aggregate of Random Variables. Correlation Coefficient

Let us examine the aggregate of two random variables ξ and η, the two-dimensional distribution function of which is equal to $\omega_2(x,y)$. Employing (2.34) and (2.66), we find the mean value of each of these variables*

$$m_1\{\xi\} = \int_{-\infty}^{\infty} x w_{11}(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x w_2(x, y)\,dx\,dy, \tag{2.86}$$

$$m_1\{\eta\} = \int_{-\infty}^{\infty} y w_{12}(y)\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y w_2(x, y)\,dx\,dy. \tag{2.86'}$$

Employing a geometrical approach, the magnitudes $m_1\{\xi\}$ and $m_1\{\eta\}$ may be regarded as the coordinates of a point which denotes the mean value of the position of a point on a plane with the random coordinates (ξ, η).

Analogously, the dispersions of each of the random variables can be found through the two-dimensional function $\omega_2(x,y)$

$$\begin{aligned} M_2\{\xi\} &= \int_{-\infty}^{\infty} (x - m_1\{\xi\})^2 w_{11}(x)\,dx = \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_1\{\xi\})^2 w_2(x, y)\,dx\,dy, \end{aligned} \tag{2.87}$$

$$\begin{aligned} M_2\{\eta\} &= \int_{-\infty}^{\infty} (y - m_1\{\eta\})^2 w_{12}(y)\,dy = \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - m_1\{\eta\})^2 w_2(x, y)\,dx\,dy. \end{aligned} \tag{2.88}$$

Let us find, for instance, the mean values and the dispersions of two normally distributed random variables. Employing (2.57) and (2.58), we obtain on the basis of the formulas

$$m_1\{\xi\} = a_1,\ m_1\{\eta\} = a_2,\ M_2\{\xi\} = \sigma_1^2,\ M_2\{\eta\} = \sigma_2^2. \tag{2.89}$$

In this manner there is determined the meaning of four (out of five) parameters of the two-dimensional distribution function (2.51), which are the mean values and the

---

* See footnote on p. 63.

dispersions of each of the two random variables subject to the normal law of distribution.

For an aggregate of two random variables, there is possible, besides the dispersion of each of the two variables, one more second moment

$$M_{12}\{\xi,\eta\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_1\{\xi\})(y-m_1\{\eta\})w_2(x,y)\,dxdy= \\ =\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyw_2(x,y)\,dxdy-m_1\{\xi\}\cdot m_1\{\eta\}, \tag{2.90}$$

which is called the mixed second moment or covariance of random variables ξ and η.

If ξ and η are independent, then $\omega_2(\underline{x},\underline{y}) = \omega_{11}(\underline{x}).\omega_{12}(\underline{y})$ and it follows from (2.90) that

$$M_{12}=\int_{-\infty}^{\infty}(x-m_1\{\xi\})w_{11}(x)\,dx\cdot\int_{-\infty}^{\infty}(y-m_1\{\eta\})w_{12}(y)\,dy=0,$$

since each of the integrals turns to zero. Therefore the values of a mixed second moment may serve as a measure of dependence between two random variables. More frequently, as such a measure there is employed the dimensionless ratio

$$R=\frac{M_{12}\{\xi\eta\}}{\sqrt{M_2\{\xi\}M_2\{\eta\}}}, \tag{2.91}$$

which is called the coefficient of correlation (coupling) [sic] between the random variables ξ and η. The magnitude of the coefficient of correlation is always contained within the limits of $-1 \leqslant \underline{R} \leqslant +1$. The limiting values are attained only when ξ and η are linearly dependent.

If the random variables are independent, $\underline{R} = 0$. The converse conclusion as to the independence of ξ and η when $\underline{R} = 0$ is in the general case invalid. Two random variables, for which the coefficient of correlation is equal to zero, are called uncorrelated. Thus, independent random variables are always uncorrelated, but not the other way around.

For instance, let $\xi_1 = \cos\xi$ and $\xi_2 = \sin\xi$, where ξ is a random variable evenly distributed between 0 and 2π. It is clear that ξ_1 and $\xi_2 = \sqrt{1-\xi_1^{\ 2}}$ are dependent, however, their covariance (and, consequently their coefficient of correla-

tion) is equal to zero*

$$m_1\{\xi_1\cdot\xi_2\}=m_1\{\cos\xi\cdot\sin\xi\}=\frac{1}{2}m_1\{\sin 2\xi\}=0.$$

Let us find the mixed second moment of two normally distributed random variables. Considering (2.89), we obtain according to (2.51) and (2.90)

$$M_{12}=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-a_1)(x_2-a_2)\exp\left\{-\frac{1}{2(1-r^2)}\times\right.$$
$$\left.\times\left[\frac{(x_1-a_1)^2}{\sigma_1^2}-2r\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2}+\frac{(x_2-a_2)^2}{\sigma_2^2}\right]\right\}dx_1dx_2.$$

Extracting the square from the expression in brackets and substituting variables

$$u=\frac{1}{\sqrt{1-r^2}}\left(\frac{x_1-a_1}{\sigma_1}-r\frac{x_2-a_2}{\sigma_2}\right),\quad v=\frac{x_2-a_2}{\sigma_2},$$

we shall have

$$M_{12}=\frac{r\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty}v^2e^{-\frac{v^2}{2}}dv\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du+$$
$$+\frac{\sigma_1\sigma_2\sqrt{1-r^2}}{2\pi}\int_{-\infty}^{\infty}ue^{-\frac{u^2}{2}}du\int_{-\infty}^{\infty}ve^{-\frac{v^2}{2}}dv,$$

from which we find

$$M_{12}=r\sigma_1\sigma_2. \tag{2.92}$$

Considering that σ_1^2 and σ_2^2 are dispersions of random variables, and keeping in mind the definition of the coefficient of correlation (2.91), we find that

$$\underline{R} = \underline{r}$$

Thus the fifth parameter $\underline{r}$ of a two-dimensional distribution function (2.51) is the coefficient of correlation between two random variables which are subject to the normal law of distribution.

In the general case, if there is an aggregate of $\underline{n}$ random variables ξ_1, ξ_2, ...ξ_n, the distribution function of which is $\omega_n(x_1, x_2, \ldots x_n)$, the mean value of the random variable ξ_k is determined by the formula

* If $m_1\{\xi_1, \xi_2\} = 0$, it is said that the random variables ξ_1 and ξ_2 are <u>orthogonal.</u> It is obvious that if the mean values of the random variables are equal to zero, the concepts of orthogonal and uncorrelated random values coincide.

$$m_1\{\xi_k\}=\underbrace{\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}}_{n\ \text{times}} x_k w_n(x_1,\ldots x_n)\,dx_1\ldots dx_n, \qquad k=1,\ 2,\ldots n. \tag{2.93}$$

The central second moments for the indicated aggregate of random variables are

$$M_{jk}\{\xi_j\xi_k\}=\underbrace{\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}}_{n\ \text{times}}(x_j-m_1\{\xi_j\})(x_k-m_1\{\xi_k\})\times$$
$$\times w_n(x_1,\ldots x_n)\,dx_1\ldots dx_n, \qquad k=1,2,\ldots n;\ j=1,2,\ldots n. \tag{2.94}$$

When $j = k$ the moment M_{kk} is equal to the dispersion of the random variable ξ_k, and when $j \neq k$, the moment $M_{jk} = M_{kj}$ is the covariance of the random variables ξ_k and ξ_j. The dimensionless ratio

$$R_{jk}=\frac{M_{jk}}{\sqrt{M_{jj}M_{kk}}} \tag{2.95}$$

is called the coefficient of correlation between the random variables ξ_k and ξ_j.

For an aggregate of n random variables there can, in addition to moments of the first and second orders, be defined moments of any order l

$$m_{k_1k_2\ldots k_n}\{\xi_1\cdot\xi_2\ldots\xi_n\}=m_1\{\xi_1^{k_1}\xi_2^{k_2}\ldots\xi_n^{k_n}\}=$$
$$=\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}}_{n\ \text{times}} x_1^{k_1}x_2^{k_2}\ldots x_n^{k_n} w_n(x_1,\ x_2,\ldots x_n)\,dx_1dx_2\ldots dx_n, \tag{2.96}$$

where $k_1 + k_2 +\ldots+ k_n\ l$ and $k_j(j = 1,2,\ldots n)$ may be any whole positive member (including also zero).

As a rule, the applications deal only with real random variables. However, for the simplification of certain theoretical investigations it is on occasion convenient to examine more general, complex random variables, which are defined by the equality $\xi = \eta + i\zeta$, where η and ζ are real random variables.

The mean value of a complex random variable is a complex number, equal to

$$m_1\{\xi\}=m_1\{\eta\}+im_1\{\zeta\}.$$

The dispersion of a complex random variable with a zero mean is equal to

$$M_2\{|\xi|\}=M_2\{\eta\}+M_2\{\zeta\}.$$

If there are given two complex random variables ξ_1 and ξ_2 with zero means, the coefficient of correlation between them is determined by the formula

$$R=\frac{M_{12}}{\sqrt{M_2^{(1)}M_2^{(2)}}},$$

where

$$M_2^{(1)}=m_1\{|\xi_1|^2\},\ M_2^{(2)}=m_1\{|\xi_2|^2\}$$

and

$$M_{12}=m_1\{\xi_1\cdot\bar{\xi}_2\}=m_1\{\eta_1\eta_2+\zeta_1\zeta_2\}+im_1\{\eta_1\zeta_2-\eta_2\zeta_1\}.$$

10. Conditional Distribution Function.

When the random variables ξ and η are dependent, there exists the conditional (a posteriori) probability that one of these random variables lies below the level of y, if the other is contained within the limits of $x < \xi \leqslant x_1$. This probability $P_{x<\xi\leqslant x_1}(\eta \leqslant y)$ is found by the multiplication rule

$$P_{x<\xi\leqslant x_1}(\eta\leqslant y)=\frac{P(x<\xi\leqslant x_1,\eta\leqslant y)}{P(x<\xi\leqslant x_1)}. \tag{2.97}$$

If there is given a two-dimensional probability density $w_2(x,y)$ of the random variables, it is possible, bearing in mind (2.28) and (2.31), to write (2.97) thus

$$P_{x<\xi\leqslant x_1}(\eta\leqslant y)=\frac{\int\limits_x^{x_1}\int\limits_{-\infty}^{y} w_2(x,y)\,dxdy}{\int\limits_x^{x_1}\int\limits_{-\infty}^{\infty} w_2(x,y)\,dxdy}. \tag{2.98}$$

Completing in (2.98) the limiting transition $x_1 \to x$ we obtain the function

$$F(y/x)=\lim_{x_1\to x}P_{x<\xi\leqslant x_1}(\eta\leqslant y)=$$

$$=\frac{\int\limits_{-\infty}^{y} w_2(x,y)\,dy}{\int\limits_{-\infty}^{\infty} w_2(x,y)\,dy}, \tag{2.99}$$

which is called the conditional integral distribution function of the random variable η, under the condition that $\xi = x$. If this function has a partial derivative

along y

$$\frac{\partial F(y/x)}{\partial y} = w(y/x), \tag{2.100}$$

this derivative is called the conditional probability density, or the conditional distribution function of the random variable η under the condition that $\xi = x$. Differentiating the right part of (2.99) with respect to y, we find

$$w(y/x) = \frac{w_2(x,y)}{\int\limits_{-\infty}^{\infty} w_2(x,y)\,dy} = \frac{w_2(x,y)}{w_{11}(x)}. \tag{2.101}$$

The conditional distribution functions have all the properties of unconditional, one-dimensional functions. In particular, it follows from (2.101) that

$$\int\limits_{-\infty}^{\infty} w(y/x)\,dy = 1. \tag{2.102}$$

Conditional numerical characteristics of a random variable may also be introduced. The conditional mean value of the random variable η, under the condition that $\xi = x$, is called the quantity

$$m_{1x}\{\eta\} = \int\limits_{-\infty}^{\infty} y w(y/x)\,dy = \frac{\int\limits_{-\infty}^{\infty} y w_2(x,y)\,dy}{\int\limits_{-\infty}^{\infty} w_2(x,y)\,dy}. \tag{2.103}$$

The conditional dispersion of the random variable η, under the condition that $\xi = x$, is equal to

$$M_{2x}\{\eta\} = \int\limits_{-\infty}^{\infty} (y - m_{1x})^2 w(y/x)\,dy = \frac{\int\limits_{-\infty}^{\infty} (y - m_{1x})^2 w_2(x,y)\,dy}{\int\limits_{-\infty}^{\infty} w_2(x,y)\,dy}. \tag{2.104}$$

The conditional distribution of the random variable ξ is determined analogously. Thus the conditional probability density under the condition $\eta = y$ is equal to

$$w(x/y) = \frac{w_2(x,y)}{\int\limits_{-\infty}^{\infty} w_2(x,y)\,dx}. \tag{2.105}$$

If the random variables ξ and η are independent, then $\omega_2(x,y) = x_{11}(x)\,\omega_{12}(y)$ and, as can be seen from (2.101) or (2.105), the conditional distribution function coincides with the unconditional distribution function of a random variable.

Let us illustrate the formulas of the present section by the example of an aggregate of two normally distributed random variables. Since each of these random variables is also distributed normally, then

$$\int_{-\infty}^{\infty} w_2(x, y)\,dy = w_{11}(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-a_1)^2}{2\sigma_1^2}},$$

and, consequently, the conditional probability density of one of the normally distributed random variables η, if the other is $\xi = x$, is according to (2.101) equal to

$$w(y/x) = \frac{1}{\sqrt{2\pi(1-r^2)}\,\sigma_2} e^{-\frac{1}{2(1-r^2)}\left\{\frac{(x-a_1)^2}{\sigma_1^2} - 2r\frac{(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right\} + \frac{(x-a_1)^2}{2\sigma_1^2}}$$

or, after reducing the similar terms in the exponent, we obtain

$$w(y/x) = \frac{1}{\sqrt{2\pi(1-r^2)}\,\sigma_2} e^{-\frac{1}{2(1-r^2)}\left(\frac{x-a_1}{\sigma_1} r - \frac{y-a_2}{\sigma_2}\right)^2}. \qquad (2.106)$$

when $r = 0$, which corresponds to the independence of the random variables, the conditional distribution function (2.106) turns to a one-dimensional function which corresponds to the normal law of distribution. The same result is obtained for $x = a_1$, but with the dispersion multiplied by $1 - r^2$.

With $r \to 1$ in accordance with (2.17)

$$w(y/x) \to \delta\left(\frac{y-a_2}{\sigma_2} - \frac{x-a_1}{\sigma_1}\right),$$

which corresponds to a linear dependence between the random variables

$$\frac{\eta - a_2}{\sigma_2} = \frac{\xi - a_1}{\sigma_1}.$$

Fig. 15 shows curves of the conditional distribution functions (2.106) for $x - a_1 + 3\sigma_1$ and for several values of the parameter r.

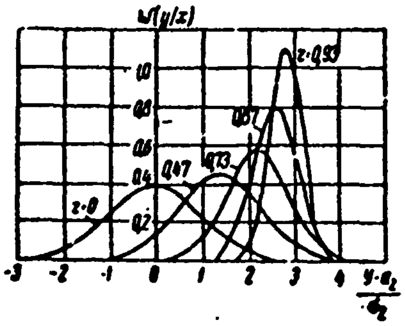


Fig. 15. Conditional distribution functions of a normal random variable under the condition that the other, dependent normal random variable takes the value of $x = a_1 + 3\sigma_1$.

Let us also find the mean value and the dispersion. According to formula (2.103), considering (2.106) we have

$$m_{1x}=\frac{1}{\sqrt{2\pi(1-r^2)}\sigma_2}\int_{-\infty}^{\infty} ye^{-\frac{1}{2(1-r^2)}\left(\frac{x-a_1}{\sigma_1}r-\frac{y-a_2}{\sigma_2}\right)^2}dy= \\ =a_2+r(x-a_1)\frac{\sigma_2}{\sigma_1}. \tag{2.107}$$

When $r = 0$ or when $x = a_1$ the conditional mean coincides with the unconditional.

When $x = a_1 + 3\sigma_1$, from (2.107) there follows $m_{1x} = a_2 + 3r\sigma_2$. This quantity determines the abscissae of the maxima of the set of conditional distribution function curves depicted in Fig. 15.

We find the conditional dispersion by formula (2.104), considering (2.106) and (2.107)

$$M_{2x}=\frac{1}{\sqrt{2\pi(1-r^2)}\sigma_2}\int_{-\infty}^{\infty}\left[y-a_2-r(x-a_1)\frac{\sigma_2}{\sigma_1}\right]^2\times \\ \times e^{-\frac{1}{2(1-r^2)}\left(\frac{x-a_1}{\sigma_1}r-\frac{y-a_2}{\sigma_2}\right)^2}dy=\sigma_2^2(1-r^2). \tag{2.108}$$

Thus the conditional dispersion does not depend on the value taken by the random variable ξ_1 but depends only on the correlation coefficient between the random variables ξ and η. When $r = 0$ the conditional dispersion coincides with the unconditional $M_{2x} = M_2 + \sigma_2^2$.

11. Measure of Uncertainty Corresponding to a Law of Distribution.

Together with the numerical characteristics (moments of distribution) examined above, it is desirable to introduce a magnitude which would characterize the degree of uncertainty created by the various laws of distribution.

In fact, every law of probability distribution describes some condition of uncertainty. Thus, a discrete random variable may take any of n values, and we know only the probabilities of the realization of these possible values. The degree of this uncertainty is different for various laws of distribution. For instance, when a random variable may take only two discrete values, the law of distribution

$$F_1(x)=\begin{cases}0,5 & x=0,\\ 0,5 & x=1\end{cases} \tag{2.109}$$

contains considerably more uncertainty than the law of distribution

$$F_2(x)=\begin{cases}0{,}9 & x=0{,}33,\\ 0{,}1 & x=2.\end{cases} \qquad (2.109')$$

It is, however, not difficult to calculate that the mean values and dispersions of two random variables, distributed according to (2.109) and (2.109'), are equal to each other. This simple example indicates that the measure of uncertainty corresponding to a given law of distribution should depend only on the magnitudes of the probabilities, and should not depend on the concrete numerical values which a random variable may take. Therefore, in introducing the indicated measure, it is necessary to digress from the numerical characteristics of random events, and to return to an examination only of their qualitative features, as was done in Ch. 1.

As a very convenient measure of uncertainty for the discrete case, there may serve the magnitude

$$H(n)=-\sum_{k=1}^{n} p_k \ln p_k, \qquad (2.110)$$

which by analogy with certain physical magnitudes is called the "entropy" of the law of distribution. In formula (2.10) p_1, p_2, ...p_n is the aggregate of the probabilities which characterize the law, under discussion, of the distribution of a discrete random variable (or the aggregate of the probabilities of a set of events). It is clear that, always,

$$\sum_{k=1}^{n} p_k = 1.$$

From the definition of entropy it follows that $H = 0$ in the case, and only in the case, that one of the numbers p_1, p_2...p_n is equal to unity (and, consequently, the others are equal to zero). But that is precisely the case when all uncertainty is absent, since it is known for certain which event must be realized. For all the other cases, when uncertainty exists, entropy is a positive magnitude.

Of all the laws of distribution of a discrete random variable, which may take any one of n possible values, the greatest uncertainty is possessed by the uniform law

$$p_k = \frac{1}{n}, \quad k = 1, 2, \ldots n.$$

For the uniform law of distribution

$$H(n) = H_{max} = -\sum_{k=1}^{n} \frac{1}{n} \ln \frac{1}{n} = \ln n. \tag{2.111}$$

Obviously, when n equiprobable results of a test are expected, the degree of uncertainty grows with the increase of n. The entropy of the uniform law of probability distribution is equal to the logarithm of this number n. Therefore it is sometimes called the logarithmic measure of uncertainty.

In the general case of n expected results of a test, some will be more probable, others less probable. However, the arbitrary law of distribution for n possible results may be replaced by an equiprobable distribution $n_1 \leqslant n$ possible results, having the same entropy

$$H(n) = \ln n_1 < \ln n.$$

For a continuous uniform distribution $\omega(x) = \frac{1}{b-a}$, $a \leqslant x \leqslant b$ the entropy must be equal to the logarithm of the width of the interval of $b - a$, i.e., $\ln(b - a)$. This magnitude for a uniform distribution coincides with $\ln \frac{1}{\omega(x)}$. It is possible to find the average statistical value of $\ln \frac{1}{\omega(x)}$ in the general case of

$$\int_{-\infty}^{\infty} \ln\left(\frac{1}{w(x)}\right) w(x)\, dx$$

and to assume it to be the measure of uncertainty of the continuous distribution (under the condition that the cited integral converges). In this manner we come to the following definition of the entropy of continuous distribution

$$H = -\int_{-\infty}^{\infty} w(x) \ln w(x)\, dx. \tag{2.112}$$

Analogously to the discrete case, of all the distribution laws of a continuous random variable, the possible values of which are bounded by the interval of (a, b), the maximum entropy $H_{max} = \ln(b - a)$ is possessed by the uniform law.

Of the aggregate of all the distribution functions which have the same dispersion

of σ^2, normal distribution is the most random, i.e., possesses the maximum entropy. The entropy of normal distribution is equal to

$$H = -\int_{-\infty}^{\infty}\left\{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}\left[\ln\frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x-a)^2}{2\sigma^2}\right]\right\}dx =$$

$$= \frac{\ln\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}}dx + \frac{1}{2\sigma^2\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}(x-a)^2 e^{-\frac{(x-a)^2}{2\sigma^2}}dx =$$

$$= \ln\sqrt{2\pi\sigma^2} + \frac{1}{2},$$

or

$$H = \ln(\sigma\sqrt{2\pi e}). \qquad (2.113)$$

Let us compare the magnitude of the entropy (2.113) of normal distribution with the magnitude of the entropy of uniform distribution, given the same dispersion of σ^2. For a dispersion of σ^2 to correspond to uniform distribution, it is necessary that the possible values of a random variable be bounded by the interval of $b - a = \sigma\sqrt{12}$ [cf. formula (3.88) in the next chapter]. The value of the entropy corresponding to this interval is equal to ln $(\sigma\sqrt{12})$. The difference between the entropy of a normal and of a uniform distribution with equal dispersions σ^2 is equal to

$$\ln(\sigma\sqrt{2\pi e}) - \ln(\sigma\sqrt{12}) = \ln\sqrt{\frac{\pi e}{6}} \approx$$

$$\approx 0.18 \text{ natural units*}$$

Let us also compute the entropy H_1 of harmonic vibration with a random phase and compare it with the entropies of normal and uniform distributions which have equal dispersions (cf. Section 2, Ch. 3)

$$H_1 = -\frac{1}{\pi}\int_{-a}^{a}\frac{1}{\sqrt{a^2-y^2}}\ln\left(\frac{1}{\pi}\cdot\frac{1}{\sqrt{a^2-y^2}}\right)dy =$$

$$= \frac{2}{\pi}\int_{0}^{a}\frac{\ln(\pi\sqrt{a^2-y^2})}{\sqrt{a^2-y^2}}dy = \ln\pi +$$

$$+ \frac{2}{\pi}\int_{0}^{\frac{\pi}{2}}\ln(a\cos\varphi)\,d\varphi = \ln\pi + \ln a - \ln 2 = \ln\frac{\pi a}{2}.$$

* More frequently the base of logarithms is taken as equal to 2 and, consequently, entropy is measured by binary units. The conversion from natural units to binary units is made by the formula

$$\log_2 M = \frac{\ln M}{\ln 2} = 1{,}43\ln M.$$

Assuming now $\frac{a^2}{2} = \sigma^2$, let us find the difference between the entropy of a normal distribution with a dispersion of σ^2 and H_1, and also between the entropy of uniform distribution with the same dispersion and H_1

$$\ln(\sigma\sqrt{2\pi e}) - H_1 = \ln(\sigma\sqrt{2\pi e}) - \ln\left(\frac{\pi}{2}\sigma\sqrt{2}\right) =$$
$$= \ln\left(2\sqrt{\frac{e}{\pi}}\right) = 0.94 \text{ natural units}$$
$$\ln(\sigma\sqrt{12}) - H_1 = \ln(\sigma\sqrt{12} - \ln\left(\frac{\pi}{2}\sigma\sqrt{2}\right) =$$
$$= \ln\left(\frac{2\sqrt{6}}{\pi}\right) = 0.76 \text{ natural units}$$

The concept of entropy can be expanded to the case of multidimensional laws of distribution. If the aggregate of random variables $\xi_1, \xi_2, \ldots \xi_n$ is linked by the n-dimensional distribution function $\omega_n(x_1, x_2, \ldots x_n)$, the entropy of this distribution is

$$H = -\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} w_n(x_1, \ldots x_n)\ln w_n(x_1, \ldots x_n)\,dx_1\ldots dx_n. \tag{2.114}$$

When $n = 2$

$$H = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_2(x, y)\ln w_2(x, y)\,dxdy. \tag{2.114'}$$

If ξ_1 and ξ_2 are independent, then by virtue of (2.36) $\omega_2(x,y)$ $\omega_{11}(x)\,\omega_{12}(y)$ and

$$\begin{aligned} H &= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_{11}(x)\,w_{12}(y)\,[\ln w_{11}(x) + \ln w_{12}(y)]\,dxdy = \\ &= -\int_{-\infty}^{\infty} w_{11}(x)\ln w_{11}(x)\,dx - \int_{-\infty}^{\infty} w_{12}(y)\ln w_{12}(y)\,dy = \\ &= H_x + H_y \end{aligned} \tag{2.115}$$

where by H_x and H_y are designated the entropies of the distribution functions of the independent random variables ξ_1 and ξ_2.

Thus the entropy of the joint distribution of two independent random variables is equal to the sum of entropies of the distribution of the component variables.

If ξ_1 and ξ_2 are dependent, then, introducing a conditional distribution function according to (2.101), we obtain

$$H=-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(y/x)\,w_{11}(x)\,[\ln w_{11}(x)+\ln w(y/x)]\,dx\,dy$$

and, considering (2.102), we find that

$$H=H_x-\int_{-\infty}^{\infty} w(y/x)\ln w(y/x)\,dy.$$

The integral

$$H_{y/x}=-\int_{-\infty}^{\infty} w(y/x)\ln w(y/x)\,dy \tag{2.116}$$

is called the conditional entropy of the random variable ξ_2, under the condition that $\xi_1 = \underline{x}$. Employing the designation of (2.116), it is possible to represent the entropy of the joint distribution of two dependent random variables in the form of

$$H=H_x+H_{y/x}. \tag{2.117}$$

If ξ_1 and ξ_2 are independent, then $\omega(\underline{y}/\underline{x})\equiv\omega_{12}(\underline{y})$ and formula (2.117) turns to formula (2.115).

Let us determine the entropy of the two-dimensional normal law of distribution.

The magnitude $\underline{H}_{\underline{x}}$ is according to (2.113) equal to $\underline{H}_{\underline{x}} = \ln(\sigma_1\sqrt{2\pi e})$, and $H_{y/x}$ we find from (2.116), utilizing (2.106)

$$H_{y/x}=-\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)\sigma_2^2}}\,e^{-\frac{1}{2(1-r^2)}\left(\frac{x-a_1}{\sigma_1}r-\frac{y-a_2}{\sigma_2}\right)^2}\times$$

$$\times\left\{\ln\frac{1}{\sqrt{2\pi(1-r^2)\sigma_2^2}}-\frac{1}{2(1-r^2)}\left(\frac{x-a_1}{\sigma_1}r-\frac{y-a_2}{\sigma_2}\right)^2\right\}dy=$$

$$=\ln\sqrt{2\pi(1-r^2)\sigma_2^2}+\frac{1}{2},$$

or

$$H_{y/x}=\ln\left(\sigma_2\sqrt{2\pi e(1-r^2)}\right). \tag{2.118}$$

Then the entropy of a two-dimensional normal distribution is

$$H=\ln(\sigma_1\sqrt{2\pi e})+\ln(\sigma_2\sqrt{2\pi e(1-r^2)})=\ln(2\pi e\sigma_1\sigma_2\sqrt{1-r^2}),$$

and, considering (2.108), we obtain

$$H=\ln(2\pi e\sigma_1\sigma_{2_1}). \tag{2.119}$$

where σ_{2x}^{2} is the conditional dispersion of ξ_2, which in the case at hand does not depend on what value is taken by ξ_1, but depends only on the coefficient of correlation between ξ_1 and ξ_2.

It can be seen from (2.118) that $H_{y/x} \leq \ln(\sigma_2\sqrt{2\pi e}) = H_y$; the equality sign corresponds to $r = 0$, i.e. when ξ_1 and ξ_2 are uncorrelated. The inequality $H_{y/x} \leq H_y$, established here for a special case, always holds true. Therefore it follows from (2.117) that

$$H \leq H_x + H_y, \tag{2.120}$$

equality being attained only for uncorrelated (and, consequently, also for independent) random variables.

Literature

1. B. V. G. Gnedenko. Kurs teorii veroyatnostey (Course in Probability Theory). Gostekhizdat, 1954.

2. G. P. Boyev. Teoriya veroyatnostey (Probability Theory). Gostekhizdat, 1950.

3. G. Kramer [i. e., Harold Cramer]. Matematicheskiye metody statistiki (Mathematical Methods of Statistics), For. Lit. Pub. Hse., 1948 (Princeton, 1946).

4. A. Ya. Khinchin. Ponyatiye entropii v teorii veroyatnostey (The Concept of Entropy in Probability Theory). "Uspekhi matematicheskikh nauk" (Progress in the Mathematical Sciences), V. VII, No.3, 1953.

Chapter III

FUNCTIONS OF CONTINUOUS RANDOM VARIABLES

1. Functional Transformations of Random Variables

In practical problems it is very often necessary, on the basis of a given distribution function $\omega_n(x_1, x_2, \ldots x_n)$ of an aggregate of random variables $\xi_1, \xi_2, \ldots \xi_n$, to determine the distribution function $W_n(y_1, y_2, \ldots y_n)$ of another aggregate of random variables $\eta_1, \eta_2, \ldots \eta_n$, which is obtained from the first by the functional transformation

$$\begin{aligned} \eta_1 &= f_1(\xi_1, \xi_2, \ldots \xi_n), \\ \eta_2 &= f_2(\xi_1, \xi_2, \ldots \xi_n), \\ &\cdots\cdots\cdots\cdots \\ \eta_n &= f_n(\xi_1, \xi_2, \ldots \xi_n), \end{aligned} \tag{3.1}$$

where $f_1, f_2, \ldots f_n$ are the given, single-valued and continuous functions.

Let us first examine the one-dimensional case. Let there be given the distribution function $\omega(x)$ of the random variable ξ; it is required to find the distribution function $W(y)$ of the random variable $\eta = f(\xi)$. Let us assume that there exists the inverse function $\xi = \varphi(\eta)$, i.e., the random variables ξ and η are linked to each other by a mutually single-valued relationship. Then from the fact that

$$x_0 < \xi \leqslant x_0 + dx, \tag{3.2}$$

it follows for certain that

$$y_0 < \eta \leqslant y_0 + dy, \quad y_0 = f(x_0) \tag{3.3}$$

and, conversely, from (3.3) there follows (3.2). Therefore the realization probabilities of these inequalities are equal to each other. The probability of the realization of inequality (3.2) is equal to the area of S_x (Fig. 16), and the probability of the realization of inequality (3.3) is equal to the area S_y.

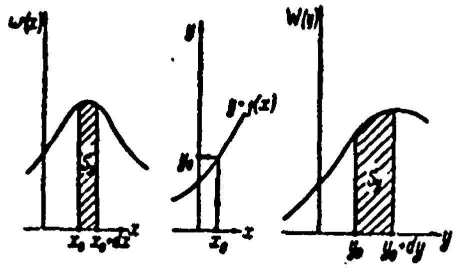


Fig. 16. Functional transformation of a random variable.

With sufficiently small values of dx and dy,

$$S_x = \omega(x)dx, \quad S_y = W(y)dy,$$

and since these areas must equal one another, we find that

$$W(y) = w(x) \cdot \frac{1}{\frac{dy}{dx}}, \qquad (3.4)$$

i.e., the desired distribution function of the random variable η, expressed in terms of the distribution function of the random variable ξ and the derivative of the well-known function $y = f(x)$. Since $W(y) \geqslant 0$, $\omega(x) \geqslant 0$, it is always necessary to substitute into formula (3.4) the absolute value of the derivative $\frac{dy}{dx}$.

If the function $y = f(x)$ is such that its inverse function $x = \varphi(y)$ is not single-valued, one value of y corresponds to several branches of the function $\varphi(y)$. Let us denote them by $x_1(y)$, $x_2(y)$...

Then from the fact that

$$y_0 < \eta \leq y_0 + dy, \qquad (3.5)$$

there follows one of the mutually exclusive possibilities

$$x_1 < \xi \leq x_1 + dx_1, \quad \text{or} \quad x_2 < \xi \leq x_2 + dx_2, \quad \text{or} \quad \ldots \qquad (3.6)$$

Employing the rule of addition, we find that the realization probability of inequality (3.5) must equal the sum of the realization probabilities of each of the inequalities (3.6). Each of these probabilities is equal respectively to the areas of S_y and S_{x_1}, S_{x_2},...(Fig. 17) and, consequently,

$$S_y = S_{x_1} + S_{x_2} + \ldots$$

or $W(y)dy = \omega(x_1)dx_1 + \omega(x_2)dx_2 + \ldots$. From the last equality we obtain the desired formula for the distribution function $W(y)$ when $x = \varphi(y)$ is multivalued

$$W(y) = w(x_1)\left|\frac{dx_1}{dy}\right| + w(x_2)\left|\frac{dx_2}{dy}\right| + \ldots \tag{3.7}$$

In view of the considerations expressed above, the absolute values of the derivative are invariably employed in (3.7)

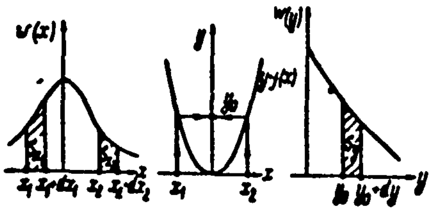


Fig. 17. Two-valued inverse transformation of random variable

It is not difficult to expand formula (3.7) to the two-dimensional case when it is necessary, on the basis of the two-dimensional distribution function $\omega_2(x_1, x_2)$ of the random variables ξ_1 and ξ_2, to determine the two-dimensional distribution function of the random variables η_1 and η_2 which are linked to the former by the definite functional relationship

$$\begin{aligned} \eta_1 &= f_1(\xi_1, \xi_2), \\ \eta_2 &= f_2(\xi_1, \xi_2). \end{aligned} \tag{3.8}$$

If the inverse transformation

$$\begin{aligned} \xi_1 &= \varphi_1(\eta_1, \eta_2), \\ \xi_2 &= \varphi_2(\eta_1, \eta_2) \end{aligned} \tag{3.9}$$

is not single-valued and has several branches $\varphi_{11}, \varphi_{21}, \varphi_{12}, \varphi_{22}, \ldots$, from the fact that point A with the coordinates η_1, η_2 is situated in some region dS (Fig. 18), it follows that point B with the coordinates ξ_1, ξ_2 is situated either in the region ds_1, or in the region ds_2, or.... Employing the rule of addition, we find that the probability of finding point A in the region dS is equal to the sum of the probabilities of finding point B in each of the regions ds_k ($k = 1,2,\ldots$). Each of these probabilities is equal respectively to the volumes V_y and $V_{x_1}, V_{x_2}, \ldots$ (Fig. 18), and, consequently,

$$\underline{V}_{\underline{Y}} = \underline{V}_{\underline{x}_1} + \underline{V}_{\underline{x}_2} + \dots$$

or

$$W(y_1, y_2)\,dS = w_2(x_{11}, x_{21})\,ds_1 + w_2(x_{12}, x_{22})\,ds_2 + \dots$$

It is well known that in the transition from the variables ($\underline{x}_1$, $\underline{x}_2$) to the variables ($\underline{y}_1$, $\underline{y}_2$), the ratio of the elementary areas $\underline{ds}$ and $\underline{dS}$ is equal to the so-called Jacobian transformation whereby

$$\frac{ds}{dS} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}, \text{ whereupon } \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \frac{1}{\frac{\partial(y_1, y_2)}{\partial(x_1, x_2)}}.$$

Consequently,

$$W_2(y_1, y_2) = w_2(x_{11}, x_{21})\left|\frac{\partial(x_{11}, x_{21})}{\partial(y_1, y_2)}\right| + \\ + w_2(x_{12}, x_{22})\left|\frac{\partial(x_{12}, x_{22})}{\partial(y_1, y_2)}\right| + \dots \tag{3.10}$$

If the correspondence between (ξ_1, ξ_2) and (η_1, η_2) is mutually single valued, only the first term remains of equation (3.10).

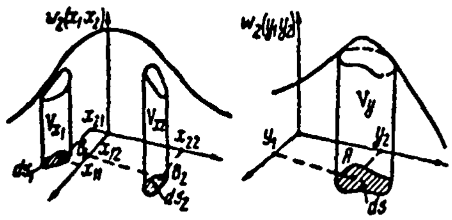


Fig. 18. Functional transformation of two random variables.

In the general case, if the Jacobian of the transformation from the random variables $\xi_1, \xi_2, \dots, \xi_{\underline{n}}$ to the random variables $\eta_1, \eta_2, \dots \eta_{\underline{n}}$ is known,

$$D = \frac{\partial(x_1, \dots x_n)}{\partial(y_1, \dots y_n)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

and if this transformation is mutually single-valued, the distribution function for $\eta_1, \eta_2, \dots \eta_{\underline{n}}$ has the form of

$$W_n(y_1, y_2, \dots y_n) = |D|\, w_n(x_1, x_2, \dots x_n). \tag{3.11}$$

2. Elementary Transformations of One Random Variable

Let us cite several examples of the employment of formula (3.7) in cases where the transformation $\eta = f(\xi)$ is performed by means of the simplest elementary functions.

The linear transformation of a random variable $\eta = a\xi + b$ is mutually single-valued. Therefore in accordance with (3.4) we have

$$W(y) = \frac{1}{|a|} w(x) = \frac{1}{|a|} w\left(\frac{y-b}{a}\right). \tag{3.12}$$

Thus in the linear transformation of a random variable its distribution curve is displaced by the amount b, and the scales of the coordinate axes change a times.

In the quadratic transformation of a random variable $\eta = \xi^2$, every value of η, which is always positive, corresponds to two values of the random variable ξ

$$\xi_1 = \sqrt{\eta}, \quad \xi_2 = -\sqrt{\eta}.$$

Then according to formula (3.7) we find, when $y > 0$,

$$W(y) = \frac{1}{2\sqrt{y}} w(\sqrt{y}) + \frac{1}{2\sqrt{y}} w(-\sqrt{y}).$$

And so the distribution function of the square of a random variable has the form of

$$W(y) = \frac{1}{\sqrt{y}} \frac{w(\sqrt{y}) + w(-\sqrt{y})}{2} \quad \text{when } y > 0, \tag{3.13}$$
$$W(y) = 0 \quad \text{when } y < 0.$$

Let, for instance, the random variable ξ be normally distributed with a mean value of zero. After quadratic transformation, the distribution function will, in accordance with (3.13), be equal to

$$W(y) = \frac{1}{2\sqrt{y}}\left(\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{y}{2\sigma^2}} + \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{y}{2\sigma^2}}\right) = \tag{3.14}$$
$$= \frac{1}{\sqrt{2\pi y}\,\sigma} e^{-\frac{y}{2\sigma^2}}, \quad y > 0$$
$$W(y) = 0, \; y < 0.$$

Let us, finally, examine the transformation of a random variable by means of the trigonometric function $\eta = a \sin \omega \xi$. With such a transformation, to each of the

possible values of η contained within the limits of from $-a$ to $+a$ there corresponds an infinite number of values of ξ

$$\xi_k = \frac{\pi k}{\omega} + \frac{(-1)^k}{\omega} \arcsin \frac{\eta}{a}, \quad k = 0, \pm 1, \pm 2, \ldots$$

According to formula (3.7) we find

$$\begin{aligned} W(y) &= \frac{1}{\omega a \sqrt{1 - \left(\frac{y}{a}\right)^2}} \sum_{k=-\infty}^{\infty} w\left[\frac{\pi k}{\omega} + \frac{(-1)^k}{\omega} \arcsin \frac{y}{a}\right], \ |y| < a \\ W(y) &= 0, \ |y| > a. \end{aligned} \tag{3.15}$$

Let, for instance, the random variable ξ be uniformly distributed over the interval of $-\frac{\pi}{\omega} \leqslant x \leqslant \frac{\pi}{\omega}$, i.e., $\omega(x) = \frac{\omega}{2\pi}$ when $|x| \leqslant \frac{\pi}{\omega}$ and $\omega(x) \equiv 0$ when $|x| \geqslant \frac{\pi}{\omega}$. Then in the infinite summation (3.15), when $k = 0$ and $k = 1$ only two terms will differ from zero.

The distribution function of the random variable $\eta = a \sin \omega \xi$ will take the following form (Fig. 19):

$$\begin{aligned} W(y) &= \frac{1}{\pi a \sqrt{1 - \left(\frac{y}{a}\right)^2}}, \ |y| < a \\ W(y) &= 0, \ |y| > a. \end{aligned} \tag{3.16}$$

Function (3.16) may be regarded as the distribution function of the values of a random-phase sinusoid, i.e., the position η of a point, moving in accordance with harmonic law in the random moment of time ξ under the condition that any of the moments within the limits of one period $T = \frac{2\pi}{\omega}$ are equiprobable. Let us note that this distribution depends only on the amplitude a of the harmonic vibrations of the point and does not depend on the period T.

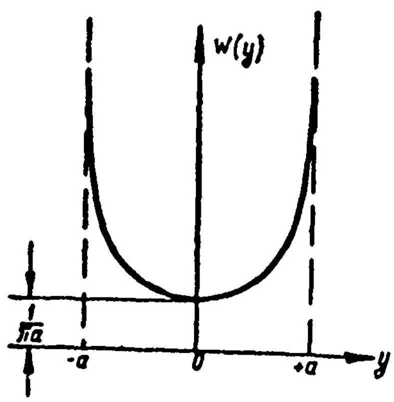


Fig. 19. Probability density of values of a sinusoid with a random phase.

Bearing in mind the relationship (2.34), it is now possible to find the distribution function of the one random variable η_2:

$$W_1(y_2) = \int_{-\infty}^{\infty} W_2(y_1, y_2)\,dy_1. \tag{3.21}$$

Substituting (3.20) into (3.21) and performing the replacement $u = \varphi_1(y_1)$, we obtain

$$W_1(y_2) = \int_{-\infty}^{\infty} \frac{w_2\{u, \varphi_2[f_1(u), y_2]\}}{\left|\frac{\partial}{\partial x_2} f_2\{u, \varphi_2[f_1(u), y_2]\}\right|}\,du. \tag{3.22}$$

Let $\eta_1 = \xi$, i.e., $f_1(x) \equiv x$; then it follows from (3.22) that

$$W_1(y_2) = \int_{-\infty}^{\infty} \frac{w_2[u, \varphi_2(u, y_2)]}{\left|\frac{\partial}{\partial x_2} f_2[u, \varphi_2(u, y_2)]\right|}\,du. \tag{3.23}$$

Formula (3.23) makes it possible to determine the distribution function $W_1(y_2)$ of the random variable η_2 obtained as the result of a functional transformation of the two random variables ξ_1 and ξ_2, the joint density of whose distribution is equal to $\omega_2(x_1, x_2)$. From this formula there are obtained as special cases, the distribution functions of the sum, difference, product and quotient of the two random variables.

Thus, for $\eta_2 = \xi_1 + \xi_2$ we have $\frac{\partial f_2}{\partial x_2} \equiv 1$, $x_2 = y_2 - x_1$ and from (3.23) we obtain the distribution function of the sum of two random variables

$$W_1(y_2) = \int_{-\infty}^{\infty} w_2(u, y_2 - u)\,du. \tag{3.24}$$

Analogously for the difference of two random variables

$$W_1(y_2) = \int_{-\infty}^{\infty} w_2(u, y_2 + u)\,du. \tag{3.24'}$$

If $\eta_2 = \xi_1 \cdot \xi_2$, then $\frac{\partial f_2}{\partial x_2} = x_1$, $x_2 = \frac{y_2}{x_1}$, and from (3.23) we obtain the distribution function of the product of two random variables

$$W_1(y_2) = \int_{-\infty}^{\infty} w_2\left(u, \frac{y_2}{u}\right)\frac{du}{|u|}. \tag{3.25}$$

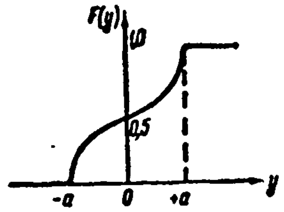


Fig. 20. Integral distribution function of values of a sinusoid with a random phase.

The integral distribution function corresponding to (3.16) is equal (Fig. 20) to

$$F(y)=\frac{1}{2}+\frac{1}{\pi}\arcsin\frac{y}{a},\quad |y|<a,$$
$$F(y)=0,\quad |y|>a. \tag{3.17}$$

4. Arithmetical Operations on Two Random Variables.

Let us examine the following special case of transformation (3.8):

$$\eta_1=f_1(\xi_1),$$
$$\eta_2=f_2(\xi_1,\ \xi_2) \tag{3.18}$$

under the condition that the reciprocal functions $\xi_1 = \varphi_1(\eta_1), \xi_2 = \varphi_2^*(\eta_2, \xi_2) = \varphi_2(\eta_1, \eta_2)$ are single-valued. The Jacobian of this transformation is equal to

$$\frac{\partial(y_1,\ y_2)}{\partial(x_1,\ x_2)}=\begin{vmatrix}\frac{df_1}{dx_1} & 0\\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2}\end{vmatrix}=\frac{df_1}{dx_1}\cdot\frac{\partial f_2}{\partial x_2}. \tag{3.19}$$

The two-dimensional distribution function of the random variables η_1 and η_2 is in accordance with (3.10) equal to

$$W_2(y_1,\ y_2)=\frac{1}{\left|\frac{d}{dx_1}f_1[\varphi_1(y_1)]\right|}\times$$
$$\times\frac{1}{\left|\frac{\partial}{\partial x_2}f_2[\varphi_1(y_1),\ \varphi_2(y_1,\ y_2)]\right|}\,w_2[\varphi_1(y_1),\ \varphi_2(y_1,\ y_2)]. \tag{3.20}$$

Analogously, the quotient of dividing one random variable by another distribution function is equal to

$$W_1(y_2) = \int_{-\infty}^{\infty} w_2(u, uy_2) \cdot |u| \, du. \tag{3.25'}$$

For independent random variables, in place of $\omega_2[\mu, \varphi_2(\mu, y_2)]$ under the sign of the integral in formulas (3.24) and (3.25), there should be placed the product $\omega_{11}(\mu)\,\omega_{12}[\varphi_2(\mu, y_2)]$ of the distribution functions of each of the random variables. Thus, the distribution function of the sum of independent random variables is equal to

$$W_1(y_2) = \int_{-\infty}^{\infty} w_{11}(u)\, w_{12}(y_2 - u)\, du. \tag{3.26}$$

The integral in the right part of (3.26), is called the <u>convolution of functions</u> ω_{11} and ω_{12}.

Let us, for instance, examine what form will be taken by the distribution function of the sum of a normally distributed random variable with a zero mean, and the value of a sinusoid with a random phase. For this it is necessary, in accordance with (3.26), to form a convolution of the distribution functions (2.14) and (3.16):

$$W_1(y) = \frac{1}{\pi a} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-a}^{a} e^{-\frac{(y-u)^2}{2\sigma^2}} \frac{du}{\sqrt{1 - \frac{u^2}{a^2}}} =$$
$$= \frac{1}{\pi\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} \int_{-1}^{+1} e^{-\frac{a^2v^2}{2\sigma^2}} e^{\frac{yav}{\sigma^2}} \frac{dv}{\sqrt{1 - v^2}}.$$

Computation of the integral leads to an expression in the form of a sum of the products of Bessel functions*

$$W_1(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2 - 2ay}{2\sigma^2}} \left[I_0\left(\frac{a^2}{4\sigma^2}\right) I_0\left(\frac{ya}{\sigma^2}\right) + \right.$$
$$\left. + 2\sum_{n=1}^{\infty} (-1)^n I_n\left(\frac{a^2}{4\sigma^2}\right) I_{2n}\left(\frac{ya}{\sigma^2}\right) \right] e^{-\frac{a^2}{4\sigma^2}} e^{-\frac{ya}{\sigma^2}}.$$

* Footnote, see p. 95

The curve of this distribution for $\left(\frac{a}{\sigma}\right)^2 = \frac{1}{5}$ is shown in Fig. 21.

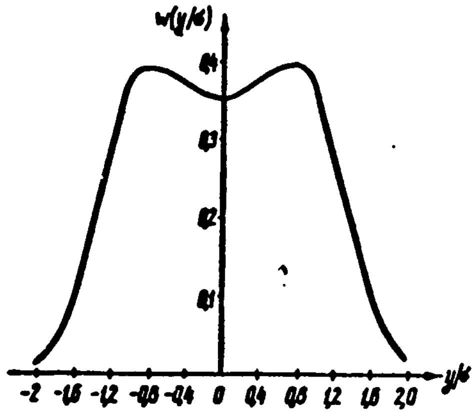


Fig. 21. Distribution of the sum of a normal random variable and a sinusoid with a random phase.

As a second example we shall examine the distribution function of two independent random variables, uniformly distributed along the segment (a,b). The probability density of each of the random variables is equal to $\frac{1}{b - a}$, if $a \leqslant x \leqslant b$, and is equal to zero outside the indicated segment. Employing (3.25), we find the desired distribution function of the product (Fig. 22):

$$W_1(y)=\int_a^{\frac{y}{a}} \frac{1}{b-a}\cdot\frac{1}{b-a}\frac{du}{u}, \quad \frac{y}{a} \geqslant a, \quad \frac{y}{b} < a,$$

$$W_1(y)=\int_{\frac{y}{b}}^{b} \frac{1}{b-a}\cdot\frac{1}{b-a}\frac{du}{u}, \quad \frac{y}{b} < b, \quad \frac{y}{a} > b,$$

$$W_1(y)=0, \quad y<a^2, \quad y>b^2$$

---

* For this it sufficient to substitute the integration variable, v = sin t, and then to employ the expansions known from the theory of Bessel functions

$$e^{b\sin t} = I_0(b) + 2\sum_{m=1}^{\infty} I_m(b)\cos m\left(\frac{\pi}{2}-t\right),$$

$$e^{c\cos t} = I_0(c) + 2\sum_{m=1}^{\infty} I_m(c)\cos mt.$$

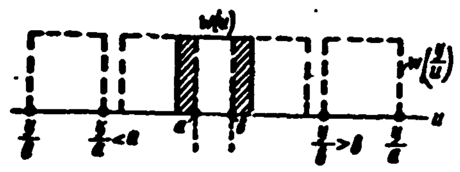


Fig. 22. Computation of the distribution function of the product of two independent, uniformly distributed random variables.

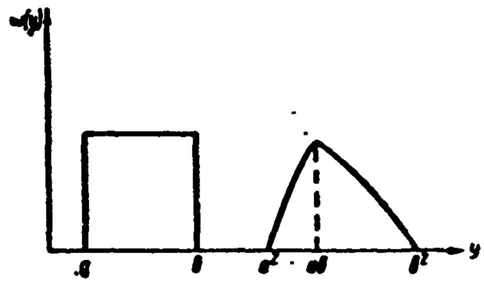


Fig. 23. Uniform distribution and distribution of the product of two independent, uniformly distributed random variables.

or, after computing the integrals, we obtain (Fig. 23)

$$W_1(y)=\frac{1}{(b-a)^2}\ln\frac{y}{a^2},\quad a^2\leqslant y\leqslant ab,$$
$$W_1(y)=\frac{1}{(b-a)^2}\ln\frac{b^2}{y},\quad ab\leqslant y\leqslant b^2,$$
$$W_1(y)=0,\quad y<a^2,\ y>b^2.$$

4. Transformation to Polar Coordinates

Let us examine one more special case of transformation (3.8):

$$\eta_1=\sqrt{\xi_1^2+\xi_2^2},\qquad (3.27)$$
$$\eta_2=\operatorname{arctg}\frac{\xi_2}{\xi_1}.$$

In (3.27) the primary value of the arc tangent is employed. This transformation is mutually single-valued, with only the positive values of η_1 possible, the possible values of the random variable η_2 being confined within the limits of from 0 to 2π. Geometrically, transformation (3.27) signifies a transition from the random cartesian coordinates (ξ_1, ξ_2) of a point, to its random polar coordinates: the length of the radius vector η_1, and the angle of inclination of this vector to the abscissa.

The transformation inverse to (3.27) has the form of

$$\begin{aligned} \xi_1 &= \eta_1 \cos \eta_2, \\ \xi_2 &= \eta_1 \sin \eta_2. \end{aligned} \tag{3.28}$$

Let there be given the two-dimensional distribution function of random cartesian coordinates $w_2(x_1, x_2)$; it is required to find the joint distribution function of random polar coordinates $w_2(\rho, \varphi)$. Since the Jacobian of transformation (3.28) from the variables x_1, x_2 to the variables ρ, φ is equal to

$$\frac{\partial (x_1, x_2)}{\partial (\rho, \varphi)} = \begin{vmatrix} \cos\varphi & -\rho\sin\varphi \\ \sin\varphi & \rho\cos\varphi \end{vmatrix} = \rho,$$

by employing formula (3.10) we obtain

$$\begin{gathered} W_2(\rho, \varphi) = \rho w_2(x_1, x_2) = \rho w_2(\rho\cos\varphi, \rho\sin\varphi) \\ \rho > 0, \ 0 \leqslant \varphi \leqslant 2\pi. \end{gathered} \tag{3.29}$$

From (3.29) we also find the one-dimensional distribution functions of the length η_1 and angle of inclination η_2 of the radius vector to be

$$\begin{gathered} W_{11}(\rho) = \rho \int_0^{2\pi} w_2(\rho\cos\varphi, \rho\sin\varphi)\, d\varphi, \ \rho > 0, \\ W_{11}(\rho) = 0, \ \rho < 0; \end{gathered} \tag{3.30}$$

$$\begin{gathered} W_{12}(\varphi) = \int_0^{\infty} \rho w_2(\rho\cos\varphi, \rho\sin\varphi)\, d\rho, \ 0 \leqslant \varphi \leqslant 2\pi, \\ W_{12}(\varphi) = 0, \ \varphi < 0, \ \varphi > 2\pi. \end{gathered} \tag{3.31}$$

It is not difficult to generalize formula (3.29) for the case of two points on a plane, whose random cartesian coordinates are dependent and characterized by the four-dimensional distribution function $\omega_4(x_1, x_2, x_3, x_4)$. The transition to lengths and inclination angles of radii-vectors is accomplished by means of the transformation

$$x_1=\rho_1\cos\varphi_1;\quad x_3=\rho_2\cos\varphi_2,$$
$$x_2=\rho_1\sin\varphi_1;\quad x_4=\rho_2\sin\varphi_2. \tag{3.32}$$

The Jacobian of transformation (3.32) is equal to

$$\frac{\partial(x_1, x_2, x_3, x_4)}{\partial(\rho_1, \varphi_1, \rho_2, \varphi_2)}=\begin{vmatrix}\cos\varphi_1 & -\rho_1\sin\varphi_1 & 0 & 0\\ \sin\varphi_1 & \rho_1\cos\varphi_1 & 0 & 0\\ 0 & 0 & \cos\varphi_2 & -\rho_2\sin\varphi_2\\ 0 & 0 & \sin\varphi_2 & \rho_2\cos\varphi_2\end{vmatrix}=\rho_1\cdot\rho_2.$$

Employing formula (3.11), we find the four-dimensional distribution of lengths and inclination angles of the radii vectors of the two points

$$W_4(\rho_1, \varphi_1, \rho_2, \varphi_2)=\rho_1\rho_2 w_4(\rho_1\cos\varphi_1, \rho_1\sin\varphi_1, \rho_2\cos\varphi_2, \rho_2\sin\varphi_2), \tag{3.33}$$
$$\rho_1>0,\ \rho_2>0,\ 0\leqslant\varphi_1\leqslant 2\pi,\ 0\leqslant\varphi_2\leqslant 2\pi.$$

From (3.33) it is not difficult to obtain the two-dimensional distribution function of the lengths $W_{21}(\rho_1, \rho_2)$ and of the angles of inclination $W_{22}(\varphi_1, \varphi_2)$ of the indicated radii vectors

$$W_{21}(\rho_1, \rho_2)=$$
$$=\rho_1\rho_2\int_0^{2\pi}\int_0^{2\pi} w_4(\rho_1\cos\varphi_1, \rho_1\sin\varphi_1, \rho_2\cos\varphi_2, \rho_2\sin\varphi_2)\,d\varphi_1 d\varphi_2, \tag{3.34}$$

$$\rho_1>0,\ \rho_2>0.$$

$$W_{21}(\rho_1, \rho_2)=0,\ \rho_1<0,\ \rho_2<0,$$
$$W_{22}(\varphi_1, \varphi_2)=$$
$$=\int_0^{\infty}\int_0^{\infty}\rho_1\rho_2 w_4(\rho_1\cos\varphi_1, \rho_1\sin\varphi_1, \rho_2\cos\varphi_2, \rho_2\sin\varphi_2)\,d\rho_1 d\rho_2, \tag{3.35}$$

$$0\leqslant\varphi_1\leqslant 2\pi,\ 0\leqslant\varphi_2\leqslant 2\pi,$$
$$W_{22}(\varphi_1, \varphi_2)=0,\ \varphi_1<0,\ \varphi_1>2\pi,\ \varphi_2<0,\ \varphi_2>2\pi.$$

In the general case, the transformation from a 2n-dimensional distribution of the random coordinates of n points to a 2n-dimensional distribution function of the lengths and inclination angles of the radii vectors of these points is made by the formula

$$W_{2n}(\rho_1, \varphi_1, \ldots \rho_n, \varphi_n) = \rho_1 \ldots \rho_n w_{2n}(\rho_1 \cos\varphi_1, \ldots \rho_n \sin\varphi_n)$$
$$\rho_1 > 0, \ldots \rho_n > 0, \ 0 \leqslant \varphi_1 \leqslant 2\pi, \ldots 0 \leqslant \varphi_n \leqslant 2\pi, \qquad (3.36)$$

special cases of which, when n = 1 and n = 2, are formulas (3.29) and (3.33).

5. Generalized Rayleigh Distribution Function

Let us apply the results of the preceding section to find the distribution function of the length of the radius vector of a point, the coordinates of which are independent and normally distributed with the parameters (a, σ) and (b, σ) respectively.

In accordance with (3.30)

$$W(\rho) = \rho \int_0^{2\pi} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(\rho\cos\varphi - a)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(\rho\sin\varphi - b)^2}{2\sigma^2}} d\varphi, \ \rho > 0,$$

or after elementary algebraic transformations, we obtain

$$W(\rho) = \frac{\rho}{2\pi\sigma^2} e^{-\frac{\rho^2 + a^2 + b^2}{2\sigma^2}} \int_0^{2\pi} e^{\frac{\rho\sqrt{a^2+b^2}}{\sigma^2}\cos(\varphi - \theta)} d\varphi, \qquad (3.37)$$
$$\rho > 0,$$

where $\Theta = \text{arc tan}\,\frac{b}{a}$.

The integral in the right side of (3.37) is, through the substitution of $\varphi - \Theta = u$ and the introduction of the designation $\alpha = \sqrt{a^2 + b^2}$, reduced to a zero-order Bessel function of an imaginary argument

$$\frac{1}{2\pi} \int_{-\theta}^{2\pi-\theta} e^{-i\left(\frac{i\alpha\rho}{\sigma^2}\right)\cos u} du = J_0\left(\frac{i\alpha\rho}{\sigma^2}\right) = I_0\left(\frac{\alpha\rho}{\sigma^2}\right).$$

Thus, the desired distribution function is equal to

$$W(\rho)=\frac{\rho}{\sigma^2}e^{-\frac{\rho^2+a^2}{2\sigma^2}}I_0\left(\frac{a\rho}{\sigma^2}\right),\ \rho>0,$$
$$W(\rho)=0,\ \rho<0. \tag{3.38}$$

Distribution function (3.38) determines the probability of finding two independent random variables, distributed according to the normal law, in the ring $(\rho, \rho + d\rho)$, whose center lies at the origin of the coordinates. This function generalizes the Rayleigh law of distribution [cf. (2.61)], which constitutes a special case of (3.38) when a = b = 0. Therefore this function may be called the generalized Rayleigh distribution function. The curves of function (3.38) for the values of $\frac{\alpha}{\sigma}$ = 0; 1; 2; 3; 5 are shown in Fig. 24. The curves of the integral distribution law corresponding to (3.38) are shown in logarithmic scale on Fig. 25. These curves were obtained through numerical integration of the generalized Rayleigh function [7].

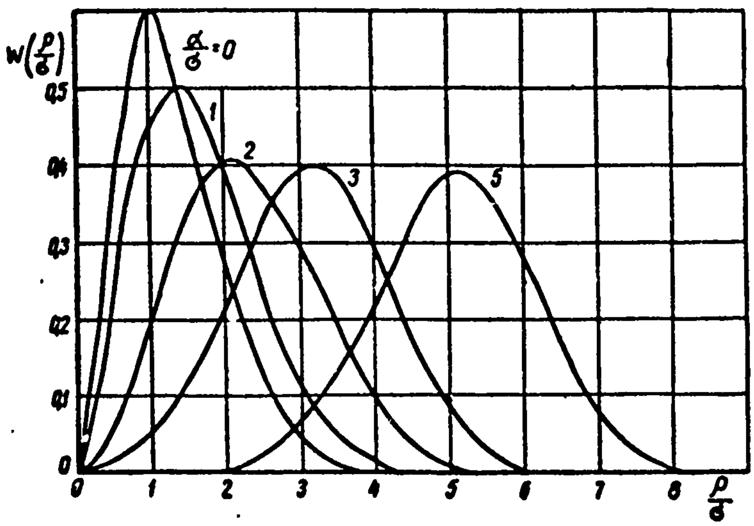


Fig. 24. Generalized Rayleigh Distribution Function

If $\frac{\alpha}{\sigma}$ is small, the generalized Rayleigh distribution function differs little from (2.61), the correction being obtainable through the expansion of the Bessel function into an exponential series. Sometimes it is possible to limit one's self to only the first two terms of this resolution and then

$$W(\rho)=\frac{\rho}{\sigma^2}e^{-\frac{\rho^2+a^2}{2\sigma^2}}\left(1+\frac{a^2\rho^2}{4\sigma^4}\right). \tag{3.39}$$

For values of $\frac{\rho}{\sigma} < \frac{\sigma}{\alpha}$ formula (3.39) is sufficiently precise, and when $\frac{\rho}{\sigma} > \frac{\sigma}{\alpha}$ the precision becomes unsatisfactory. But if $\frac{\alpha}{\sigma} \ll 1$, the probability density for $\frac{\rho}{\sigma} > \frac{\sigma}{\alpha}$ becomes so small that it is generally disregarded in practice.

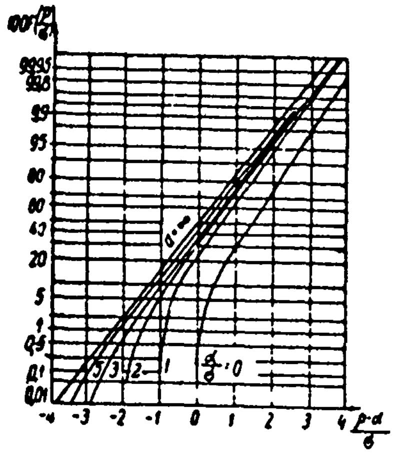


Fig. 25. Integral Generalized Rayleigh Distribution Function

If $\frac{\alpha}{\sigma}$ is large, then in expression (3.38) the Bessel function may be replaced by its asymptotic expansion

$$I_0(z) \sim \frac{e^z}{\sqrt{2\pi z}}\left(1+\frac{1}{8z}+\dots\right).$$

Then

$$W(\rho) \sim \frac{\rho}{\sigma^2} e^{-\frac{\rho^2+a^2}{2\sigma^2}} \cdot \frac{\sigma}{\sqrt{2\pi\rho a}} e^{\frac{\rho a}{\sigma^2}}\left(1+\frac{\sigma^2}{8\rho a}\right),$$

or

$$W(\rho) \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(\rho-a)^2}\left(1+\frac{\sigma^2}{8a\rho}\right)\sqrt{\frac{\rho}{a}},\ \rho>0. \qquad (3.40)$$

From formula (3.40) it follows that, with accuracy equal to the correction factor of $\left(1+\frac{\sigma^2}{8\alpha\rho}\right)\sqrt{\frac{\rho}{\alpha}}$, the generalized Rayleigh distribution law turns in this case into the normal law of distribution with the parameters of α and σ. If $\frac{\alpha}{\sigma} \gg 1$ and if $\left|\frac{\rho}{\sigma}-\frac{\alpha}{\sigma}\right|$ is not large, the correction of the normal law is insignificant, and for cases of large $\left|\frac{\rho}{\sigma}-\frac{\alpha}{\sigma}\right|$, although the correction is substantial, the probability density at these values becomes insignificantly small.

Let us determine the numerical characteristics of a random variable distributed according to the generalized Rayleigh law. In accordance with (2.64) the k-th

initial moment of distribution (3.38) is equal to

$$m_k=\frac{1}{\sigma^2}e^{-\frac{a^2}{2\sigma^2}}\int_0^\infty \rho^{k+1}I_0\left(\frac{a\rho}{\sigma^2}\right)e^{-\frac{\rho^2}{2\sigma^2}}d\rho.$$

As a result of computing the integral [cf. for instance, Vatson, G. N. Teoriya besselevykh funktsiy (Theory of Bessel Functions), Moskva, For. Lit. Pub. Hse., 1949 (i.e., Watson, G. N., "A Treatise on the Theory of Bessel Functions", N. Y. 1944)] we obtain

$$m_k=(2\sigma^2)^{\frac{k}{2}}\Gamma\left(1+\frac{k}{2}\right){}_1F_1\left(-\frac{k}{2},\ 1,\ -\frac{a^2}{2\sigma^2}\right), \tag{3.41}$$

where ${}_1F_1$–is a degenerate hypergeometric function, the basic properties of which are cited in Appendix VI.

The mean value m_1 of a random variable distributed according to the generalized Rayleigh law is

$$\begin{aligned} m_1&=\sigma\sqrt{2}\,\Gamma\left(\frac{3}{2}\right){}_1F_1\left(-\frac{1}{2},\ 1,\ -\frac{a^2}{2\sigma^2}\right)=\\ &=\sigma\sqrt{\frac{\pi}{2}}\left[\left(1+\frac{a^2}{2\sigma^2}\right)I_0\left(\frac{a^2}{4\sigma^2}\right)+\right.\\ &\left.+\frac{a^2}{2\sigma^2}I_1\left(\frac{a^2}{4\sigma^2}\right)\right]e^{-\frac{a^2}{4\sigma^2}}. \end{aligned} \tag{3.42}$$

From (3.42) as a special case with $\alpha = 0$ we obtain formula (2.79) for the mean value corresponding to the Rayleigh distribution.

Employing (3.41) it is also not difficult to write the expressions for initial moments of the 2-nd and 3-rd orders,

$$m_2=2\sigma^2+a^2, \tag{3.43}$$

$$\begin{aligned} m_3&=3\sigma^3\sqrt{\frac{\pi}{2}}\left[\left(1+\frac{a^2}{\sigma^2}+\frac{a^4}{6\sigma^4}\right)I_0\left(\frac{a^2}{4\sigma^2}\right)+\right.\\ &\left.+\left(\frac{2a^2}{3\sigma^2}+\frac{a^4}{6\sigma^4}\right)I_1\left(\frac{a^2}{4\sigma^2}\right)\right]e^{-\frac{a^2}{4\sigma^2}}, \end{aligned} \tag{3.44}$$

special cases of which, when $\alpha = 0$, are the corresponding formulas of Section 8, Ch.2. Considering (2.74), it is likewise not difficult to obtain from (3.42) and (3.43)

the expression for the dispersion of a random variable which is distributed according to the generalized Rayleigh law.

If $\alpha \gg \sigma$, then, employing the previously cited asymptotic resolution of the Bessel function, we find that

$$m_1 \sim \alpha + \frac{\sigma^2}{2\alpha} = \alpha\left(1 + \frac{\sigma^2}{2\alpha^2}\right). \tag{3.45}$$

$$M_2 \sim \sigma^2 - \frac{\sigma^4}{4\alpha^2} = \sigma^2\left(1 - \frac{\sigma^2}{4\alpha^2}\right). \tag{3.46}$$

The first items in formulas (3.45) and (3.46) are respectively the mean value and the dispersion of the limiting normal law of distribution, and the second items provide a correction which diminishes with the growth of $\frac{\alpha}{\sigma}$.

Substituting (3.42), (3.43) and (3.44) in (2.81), we find the central moment of the third order, after which, considering (2.82), it is possible to compute the coefficient of asymmetry, which will depend only on $\frac{\alpha}{\sigma}$. Results of the computation of the coefficient of asymmetry k, for the curves depicted in Fig. 24, are cited in Table 3.

Table 3

α	0	1	2	3	5
k	0.63	0.43	0.24	0.07	0.015

The obtained data make it possible to judge the extent to which the coefficient of asymmetry is sensitive to changes in the form of the curve. From a comparison of Fig. 24 with the data of Table 3, it is evident that when k = 0.07, i.e., with a diminution in the coefficient of asymmetry of about 10 times in comparison to the most asymmetrical Rayleigh distribution (α = 0), the distribution curve becomes practically symmetrical and is in sufficient proximity to the curve of the normal law of distribution.

6. Numerical Characteristics of Functions of Random Variables.

Let there be given the distribution function $\omega(x)$ of the random variable ξ ; it is required to find the mean value of the random variable $\eta = f(\xi)$. If the inverse function $\xi = \varphi(\eta)$ is single-valued, then, employing (3.4), we obtain

$$m_1\{\eta\} = \int_{-\infty}^{\infty} yW(y)\,dy = \int_{-\infty}^{\infty} f(x)\,w(x)\frac{dx}{dy}\,dy,$$

or

$$m_1\{\eta\} = \int_{-\infty}^{\infty} f(x)\,w(x)\,dx. \tag{3.47}$$

It can be demonstrated that formula (3.47) of the mean value of a function of a random variable is valid for any continuous function $f(x)$.

Let $\eta = \xi^k$. Then from (3.47) it follows that

$$m_1\{\xi^k\} = \int_{-\infty}^{\infty} x^k w(x)\,dx = m_k\{\xi\}. \tag{3.48}$$

i.e., the k-th moment of distribution may be treated as the mean value of the k-th degree of the random variable.

Analogously,

$$M_k\{\xi\} = m_k\{\xi - m_1\} = m_1\{(\xi - m_1)^k\}. \tag{3.49}$$

Since the mean value of a constant is equal to the constant itself, then from (3.49) it follows for the special case of $\xi = c$ that the central moment of a constant of any order is equal to zero.

In the general case, when there is given a distribution function of the aggregate of $\xi_1, \ldots \xi_n$ random variables and it is required to find the value of $\eta = f(\xi_1, \ldots \xi_n)$, we have

$$m_1\{\eta\} = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ times}} f(x_1, \ldots x_n)\,w_n(x_1, \ldots x_n)\,dx_1 \ldots dx_n. \tag{3.50}$$

The special case of formula (3.50) for $n = 2$ may be employed for finding the numerical characteristics of a sum and of a product of random variables.

The mean value of a sum (or difference) of random variables is

$$m_1\{\xi_1 \pm \xi_2\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 \pm x_2)\,w_2(x_1,\ x_2)\,dx_1dx_2 =$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1 w_2(x_1,\ x_2)\,dx_1dx_2 \pm$$
$$\pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_2 w_2(x_1,\ x_2)\,dx_1dx_2,$$

and, considering (2.86) and (2.86'), we find that

$$m_1\{\xi_1 \pm \xi_2\} = m_1\{\xi_1\} \pm m_1\{\xi_2\}. \tag{3.51}$$

From formula (3.51) it follows that

$$m_1\left\{\sum_{k=1}^{n}\xi_k\right\} = \sum_{k=1}^{n} m_1\{\xi_k\}. \tag{3.52}$$

Thus, the mean value of a sum of random variables is always equal to the sum of the mean values of the terms.

The mean value of the product of random variables ξ_1, ξ_2 is, according to (2.90) equal to

$$\begin{aligned} m_1\{\xi_1\cdot\xi_2\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1 x_2 w_2(x_1,\ x_2)\,dx_1dx_2 = \\ &= m_1\{\xi_1\}\cdot m_1\{\xi_2\} + M_{12}\{\xi_1\cdot\xi_2\}. \end{aligned} \tag{3.53}$$

If ξ_1 and ξ_2 are independent, then $M_{12}\{\xi_1\cdot\xi_2\} = 0$ and

$$m_1\{\xi_1\cdot\xi_2\} = m_1\{\xi_1\}\cdot m_1\{\xi_2\}. \tag{3.53'}$$

It follows from formula (3.53') that for mutually independent random variables

$$m_1\left\{\prod_{k=1}^{n}\xi_k\right\} = \prod_{k=1}^{n} m_1\{\xi_k\}. \tag{3.54}$$

Thus the mean value of a product of mutually independent random variables is equal to the product of the mean values of the factors.

The following rule is a consequence of (3.53'): the constant factor c may be brought out beyond the sign of the mean value

$$m_1\{c\xi\} = cm_1\{\xi\}. \tag{3.55}$$

Let us now examine the dispersion of the sum (or difference) of two random variables. Employing (3.49) and (3.52), we find that

$$\begin{aligned} M_2\{\xi_1 \pm \xi_2\} &= m_1\{[\xi_1 \pm \xi_2 - m_1\{\xi_1 \pm \xi_2\}]^2\} = \\ &= m_1\{[(\xi_1 - m_1\{\xi_1\}) \pm (\xi_2 - m_1\{\xi_2\})]^2\} = \\ &= M_2\{\xi_1\} + M_2\{\xi_2\} \pm 2m_1\{[\xi_1 - m_1\{\xi_1\}][\xi_2 - m_1\{\xi_2\}]\} \end{aligned}$$

and, considering (2.90), we obtain

$$M_2\{\xi_1 \pm \xi_2\} = M_2\{\xi_1\} + M_2\{\xi_2\} \pm 2M_{12}\{\xi_1 \cdot \xi_2\}. \tag{3.56}$$

If ξ_1 and ξ_2 are independent, their mixed moment $M_{12} = 0$, and then

$$M_2\{\xi_1 \pm \xi_2\} = M_2\{\xi_1\} + M_2\{\xi_2\}, \tag{3.57}$$

i.e., the dispersion of the sum or difference of two independent random variables is equal to the sum of the dispersions of these variables. From formula (3.57) it follows, that for mutually independent random variables

$$M_2\left\{\sum_{k=1}^{n} \pm \xi_k\right\} = \sum_{k=1}^{n} M_2\{\xi_k\}. \tag{3.58}$$

Thus the dispersion of an algebraic sum of mutually independent random variables is equal to the sum of the dispersions of the terms. The dispersion of a product of independent random variables is equal to

$$\begin{aligned} M_2\{\xi_1 \cdot \xi_2\} &= m_1\{[\xi_1\xi_2 - m_1\{\xi_1\} \cdot m_1\{\xi_2\}]^2\} = M_2\{\xi_1\} \cdot M_2\{\xi_2\} + \\ &+ m_1^2\{\xi_1\} M_2\{\xi_2\} + m_1^2\{\xi_2\} \cdot M_2\{\xi_1\}. \end{aligned} \tag{3.59}$$

Expressing dispersions in terms of initial moments according to formula (2.74), after elementary transformations we obtain from (3.59)

$$M_2\{\xi_1 \cdot \xi_2\} = m_2\{\xi_1\} \cdot m_2\{\xi_2\} - m_1^2\{\xi_1\} \cdot m_1^2\{\xi_2\}. \tag{3.59'}$$

Formula (3.59') generalizes formula (2.74) for the two-dimensional case.

Since the dispersion of a constant is equal to zero, from (3.59) there follows the rule: a constant factor may be brought out beyond the sign of dispersion, if this factor is squared in such a case

$$M_2\{c\xi\} = c^2 M_2\{\xi\}. \tag{3.60}$$

According to this rule, for instance, the dispersion of the normalized deviation $\frac{\xi - m_1}{\sqrt{M_2}}$ of the random variable ξ is always equal to unity.

7. The Characteristic Function

In the preceding section there has been presented the general formula (3.47) for the mean value of the random variable η, which is obtained by means of the functional transformation f of the random variable ξ, the distribution function of which is equal to $\omega(x)$. In subsequent applications an important role will be played by one special form of this transformation, namely

$$\eta = e^{iv\xi}, \tag{3.61}$$

where v is an arbitrary, real parameter. The mean value of the random variable $e^{iv\xi}$ is called the characteristic function of the random variable ξ or the characteristic function of a given probability distribution. In accordance with formula (3.47) the characteristic function of the random variable ξ is equal to

$$\theta(v) = m_1\{e^{iv\xi}\} = \int_{-\infty}^{\infty} w(x)\, e^{ivx}\, dx. \tag{3.62}$$

Since $\left| e^{ivx} \right| = 1$, the integral (3.62) converges at all real values of v for all distribution functions $\omega(x)$. Therefore a characteristic function may be determined for any random variable.

Let us formulate the basic properties of a characteristic function.

1. From the definition it follows that

$$\Theta(0)=\int_{-\infty}^{\infty} w(x)\,dx=1 \text{ и } |\Theta(v)|\leqslant\Theta(0)=1. \quad (3.63)$$

2. If $\Theta_{11}(v)$ is the characteristic function of the random variable ξ_1, the characteristic function $\Theta_{12}(v)$ of the random variable ξ_2, which is obtained by means of the linear transformation $\xi_2 = \underline{a}\,\xi_1 + \underline{b}$, is equal to

$$\Theta_{12}(v)=\int_{-\infty}^{\infty} e^{iv(ax+b)}\,w(x)\,dx=\Theta_{11}(av)\,e^{ibv}. \quad (3.64)$$

3. If there exists a $\underline{k}$-th initial moment of distribution of the random variable ξ, the characteristic function of this variable has a derivative of the $\underline{k}$-th order, with

$$\frac{d^k\Theta(v)}{dv^k}=i^k\int_{-\infty}^{\infty} x^k e^{ivx}\,w(x)\,dx,$$

wherefrom

$$\left(\frac{d^k\Theta(v)}{dv^k}\right)_{v=0}=i^k m_k\{\xi\}. \quad (3.65)$$

Thus the initial moments of distribution differ only by the factor $\underline{i}^{\underline{k}}$ from the value of the derivatives of the characteristic function when $\underline{v} = 0$. From (3.65), in the special case of $\underline{k} = 1$, we obtain the expression for the mean value

$$m_1\{\xi\}=\frac{1}{i}\Theta'(0). \quad (3.66)$$

If moments of every order exist, then, as follows from (3.65), the characteristic function may be represented by a MacLaurin series

$$\Theta(v)=1+\sum_{k=1}^{\infty}\frac{m_k}{k!}(iv)^k. \quad (3.67)$$

4. Central moments of distribution are linked by simple relationships to the derivatives of the logarithm of the characteristic function. Let us assume $\psi(\underline{v})$ =

$= \ln \Theta(v)$; then

$$\psi''(v) = \frac{\Theta''(v)\,\Theta(v) - [\Theta'(v)]^2}{\Theta^2(v)},$$

and, considering (3.63) and (3.65) we obtain

$$\psi''(0) = \Theta''(0) - [\Theta'(0)]^2 = \\ = -m_2\{\xi\} + m_1^2\{\xi\} = -M_2\{\xi\},$$

i.e., the dispersion of a random variable is equal to

$$M_2\{\xi\} = -\psi''(0). \tag{3.68}$$

Analogously it is possible to obtain the formulas

$$M_3 = -i^3\psi'''(0),\ M_4 = i^4\psi^{(IV)}(0) + 3M_2^2,$$

from which also follow the expressions for the coefficients of asymmetry and excess

$$k = \frac{\psi'''(0)}{[\psi''(0)]^{3/2}},\ \gamma = \frac{\psi^{IV}(0)}{[\psi''(0)]^2}. \tag{3.69}$$

The k-th order derivative of the logarithm of a characteristic function at the point of $v = 0$, multiplied by i^k, is called the k-th order cumulant or semi-invariant of a random variable. As can be seen from the formulas presented, a knowledge of the semi-invariants of the first four orders makes it possible to find very simply the mean value, the dispersion, and the coefficients of asymmetry and excess.

5. Let ξ_1 and ξ_2 be independent random variables, the characteristic functions of which are equal respectively to $\Theta_{11}(v)$ and $\Theta_{12}(v)$. We are to find the characteristic function $\Theta(v)$ of the sum $\eta = \xi_1 + \xi_2$ of these random variables. Since, together with ξ_1 and ξ_2, $e^{iv\xi_1}$ and $e^{iv\xi_2}$ are also independent, by employing the property (3.53')of the mean value of a product of independent random variables, we obtain

$$\Theta(v) = m_1\{e^{iv(\xi_1+\xi_2)}\} = m_1\{e^{iv\xi_1}\cdot e^{iv\xi_2}\} = \\ = m_1\{e^{iv\xi_1}\}\cdot m_1\{e^{iv\xi_2}\},$$

or

$$\Theta(v) = \Theta_{11}(v)\,\Theta_{12}(v). \tag{3.70}$$

Thus, the characteristic function of a sum of two independent random variables is equal to the product of the characteristic functions of the items.* This rule for calculating characteristic functions can be directly expanded for the case of any number of mutually independent items, i.e., if $\Theta_{1k}(v)$ is the characteristic function of the random variable ξ_k, the characteristic function of the summation $\eta = \sum_{K=1}^{n} \xi_K$ is equal to

$$\Theta(v) = \prod_{k=1}^{n} \Theta_{1k}(v). \tag{3.71}$$

In the special case when all the items have the same distribution $\Theta_{1k}(v) = \Theta_1(v)$,

$$\Theta(v) = \Theta_1^n(v). \tag{3.71'}$$

The cited properties of characteristic functions make them a particularly convenient tool for the study of sums of mutually independent random variables. While the distribution function of a sum of two independent random variables is obtained by the complex operation of convolving the distribution functions of the terms (cf. 3.26), the characteristic function of this sum is found through the simple multiplication of the characteristic functions of the terms. If the characteristic function of a random variable is known, its distribution function can always be found. The latter follows from the fact that, as is evident from (3.62), the characteristic function $\Theta(v)$ is obtained by the direct transformation of the Fourier distribution function $\omega(x)$. Therefore the distribution function is obtained from the characteristic function by means of the inverse Fourier transformation

* Let us note that in the derivation of formula (3.53') and, consequently, also of (3.70), it was sufficient to require that ξ_1 and ξ_2 be uncorrelated.

$$w(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Theta(v)\,e^{-ivx}dv. \qquad (3.72)$$

Thus, to calculate the distribution function of a sum of independent random variables it is necessary first to compute the characteristic functions of the items, and then to determine the inverse Fourier transformation of the product of these characteristic functions, which will be the desired distribution function.

6. Employing formula (3.47), it is not difficult to find the expression for the characteristic function of the random variable $\eta = f(\xi)$, if the distribution function $\omega(x)$ of the random variable ξ is known

$$\Theta(v)=m_1\{e^{ivf(\xi)}\}=\int_{-\infty}^{\infty}w(x)\,e^{ivf(x)}dx. \qquad (3.73)$$

7. The method of characteristic functions can be expanded to the aggregate of random variables $\xi_1, \xi_2, \ldots \xi_n$. The characteristic function of the aggregate of random variables $\xi_1, \xi_2, \ldots \xi_n$ is called the mean value of the random variable $e^{i(v_1\xi_1+v_2\xi_2+\ldots v_n\xi_n)}$, where $v_1, v_2, \ldots v_n$ are real parameters. If $w_n(x_1, x_2, \ldots x_n)$ is an n-dimensional distribution function of an aggregate of random variables, then

$$\Theta_n(v_1, v_2, \ldots v_n)=m_1\{e^{i(v_1\xi_1+v_2\xi_2+\ldots+v_n\xi_n)}\}=$$
$$=\underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n\text{ times}}e^{i(v_1x_1+v_2x_2+\ldots+v_nx_n)}\times \qquad (3.74)$$
$$\times w_n(x_1, x_2, \ldots x_n)\,dx_1dx_2\ldots dx_n.$$

Consequently, the n-dimensional characteristic function is the n-fold Fourier transformation of an n-dimensional distribution function. Conversely, the distribution function of an aggregate of n random variables is obtained from their characteristic function by means of an n-fold inverse Fourier transformation.

On the basis of the characteristic function $\Theta_n(v_1, v_2, \ldots v_n)$ of the aggregate of random variables $\xi_1, \xi_2, \ldots \xi_n$, it is not difficult to find the characteristic function of any aggregate of $k < n$ variables

$$\Theta_k(v_1, v_2, \ldots v_k) = \Theta_n(v_1, v_2, \ldots v_k, 0, \ldots 0). \tag{3.74'}$$

If the random variables $\xi_1, \xi_2, \ldots \xi_n$ are mutually independent, their n-dimensional characteristic function is equal to the product of the characteristic functions of each of the random variables

$$\Theta_n(v_1, v_2, \ldots v_n) = \prod_{k=1}^{n} \Theta_{1k}(v_k). \tag{3.75}$$

From (3.74) when $v_1 = v_2 = \ldots = v_n = v$ we obtain the expression for the characteristic function of the sum of $\xi_1 + \xi_2 + \ldots + \xi_n$ dependent random variables

$$\Theta(v) = \Theta_n(v, v, \ldots v). \tag{3.76}$$

Formula (3.71) is a special case of (3.76), when the component sums are mutually independent.

8. Let $\Theta_{2k}(v_1, v_2)$ be the characteristic function* of the random variables (ξ_k, η_k) and let the pairs of random variables (ξ_1, η_1), $(\xi_2, \eta_2), \ldots (\xi_n, \eta_n)$, be mutually independent. The characteristic function of the random variables

$$\xi = \sum_{k=1}^{n} \xi_k, \quad \eta = \sum_{k=1}^{n} \eta_k$$

is equal to

$$\Theta_2(v_1, v_2) = \prod_{k=1}^{n} \Theta_{2k}(v_1, v_2). \tag{3.77}$$

If the random variables ξ_k and η_k are treated as components of a plane vector (3.77) signifies that the characteristic function of a sum of independent vectors (i.e., the resultant of the vectors) is equal to the product of the characteristic functions of the component vectors. This rule remains true for vectors in multidimensional space. If all the summed-up vectors are subject to the same two-dimensional law of distribution $\Theta_2(v_1 v_2)$, then

$$\Theta_2(v_1, v_2) = [\Theta_2(v_1, v_2)]^n. \tag{3.77'}$$

9. Multidimensional characteristic functions may also be employed for determining the mixed moments of distribution of an aggregate of random variables (cf. 2.96).

* A double index is used with characteristic functions in the same sense as with distribution functions (cf. reference on p. 51).

If there exists the derivative

$$\frac{\partial^{k_1+k_2+\ldots+k_n}}{\partial v_1^{k_1}\partial v_2^{k_2}\ldots\partial v_n^{k_n}}[\Theta_n(v_1, v_2, \ldots v_n)] =$$

$$= i^{k_1+k_2+\ldots+k_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_1^{k_1}x_2^{k_2}\ldots x_n^{k_n}\times$$

$$\times w_n(x_1, x_2, \ldots x_n)\exp[i(v_1x_1+v_2x_2+\ldots v_nx_n)]\,dx_1\,dx_2\ldots dx_n,$$

then

$$m_{k_1k_2\ldots k_n}\{\xi_1, \xi_2\ldots\xi_n\} = i^{-(k_1+k_2+\ldots+k_n)}\times$$
$$\times\left\{\frac{\partial^{k_1+k_2+\ldots+k_n}}{\partial v_1^{k_1}\partial v_2^{k_2}\ldots\partial v_n^{k_n}}[\Theta_n(v_1, v_2, \ldots v_n)]\right\}_{v_1=v_2=\ldots=v_n=0}. \qquad (3.78)$$

Obviously (3.65) is a special case of (3.78).

8. Computation of Distribution Moments

Let us apply the method of characteristic functions to the calculation of distribution moments of four types: normal, Rayleigh, uniform, and the distribution of the values of a sinusoid with a random phase.

The characteristic function of a random variable, normally distributed with the parameters a, σ, is in accordance with (3.62) amd (2.14) equal to

$$\Theta(v) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{ivx}e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx.$$

The substitution of $z = \frac{x-a}{\sigma} - i\sigma v$ is used to complete the square in the exponent in the integrand function. Then

$$\Theta(v) = e^{iav-\frac{\sigma^2v^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty-i\alpha}^{\infty-i\alpha} e^{-\frac{z^2}{2}}\,dz.$$

Since with any instance of a real α

$$\int_{-\infty-i\alpha}^{\infty-i\alpha} e^{-\frac{z^2}{2}}\,dz = \sqrt{2\pi},$$

it follows that

$$\Theta(v) = e^{iav - \frac{\sigma^2 v^2}{2}}. \tag{3.79}$$

Employing (3.65), and considering that central moments coincide with the initial moments at the zero mean value ($a = 0$), we find the central moments of normal distribution to be

$$M_k = \frac{1}{i^k}\left(\frac{d^k}{dv^k} e^{-\frac{\sigma^2 v^2}{2}}\right)_{v=0}.$$

Expressing the derivatives of the function $e^{-\frac{\sigma^2 v^2}{2}}$ in terms of Hermite polynomials (cf. Appendix VII), we obtain

$$M_k = i^k \sigma^k H_k(0),$$

wherefrom

$$M_{2k} = \sigma^{2k}(2k-1)!!, \quad M_{2k+1} = 0, \tag{3.80}$$

where $(2k - 1)!!$ is the product of all the odd numbers in the natural series to $2k - 1$ inclusive.

The logarithm of the characteristic function of a normal random variable is equal to

$$\psi(v) = \ln\Theta(v) = iav - \frac{\sigma^2 v^2}{2}. \tag{3.79'}$$

It follows from (3.79') that the first-order cumulant of a normal random variable coincides with its mean value, the second-order cumulant with the dispersion, and all the cumulants of an order higher than two are equal to zero.

The characteristic function of the square of a normally-distributed random variable with a zero mean value and a dispersion of σ^2 is, in accordance with (3.14), equal to

$$\Theta(v)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_0^\infty y^{-\frac{1}{2}}e^{-\left(\frac{1}{2\sigma^2}-iv\right)y}dy=$$

$$=\frac{\Gamma\left(\frac{1}{2}\right)}{\sqrt{2\pi\left(\frac{1}{2\sigma^2}-iv\right)\sigma^2}}=(1-2i\sigma^2v)^{-\frac{1}{2}}. \tag{3.81}$$

Expanding (3.81) into a series of powers of y, we obtain

$$\Theta(v)=1+\sum_{k=1}^{\infty}\frac{(2k-1)!!}{(2k)!!}(2\sigma^2)^k(iv)^k, \tag{3.81'}$$

where (2k)!! is the product of all the even numbers of the natural series to 2k inclusive.

Comparing (3.81') with (3.67), we find the distribution moments of the square of a normally distributed random variable to be

$$m_k=(2\sigma^2)^k\frac{(2k-1)!!\,k!}{(2k)!!}. \tag{3.82}$$

The characteristic function of a random variable, distributed according to the Rayleigh law in accordance with (3.62) and (2.61) is equal to

$$\Theta(v)=\frac{1}{\sigma^2}\int_0^\infty xe^{ivx-\frac{x^2}{2\sigma^2}}dx=1+i\sigma v\sqrt{\frac{\pi}{2}}\left[1+\Phi\left(\frac{i\sigma v}{\sqrt{2}}\right)\right]e^{-\frac{\sigma^2v^2}{2}} \tag{3.83}$$

where Φ is the Kramp function.

Employing (3.83) and considering (3.65), by successive differentiation we find the Rayleigh distribution moments

$$m_1=\sqrt{\frac{\pi}{2}}\sigma,\ m_2=2\sigma^2,\ m_3=3\sqrt{\frac{\pi}{2}}\sigma^3,\ m_4=8\sigma^4,$$

or in a general form

$$m_k=(2\sigma^2)^{\frac{k}{2}}\Gamma\left(\frac{k}{2}+1\right). \tag{3.84}$$

Let us examine the characteristic function of a random variable uniformly dis-

tributed over the interval of (a, b). Since the probability density within the limits of the indicated interval is constant and equal to $\frac{1}{b - a}$, and outside of this interval is equal to zero, it follows that

$$\Theta(v) = \frac{1}{b-a}\int_a^b e^{ivx}dx = \frac{1}{iv}\cdot\frac{e^{ibv} - e^{iav}}{b-a}. \tag{3.85}$$

The distribution moments in this case are easiest of all to find, if (3.85) is resolved into the MacLaurin series

$$\Theta(v) = \frac{1}{b-a}\sum_{k=0}^{\infty}\frac{(iv)^k}{k!(k+1)}(b^{k+1} - a^{k+1}) \tag{3.86}$$

and the coefficients for like powers of v are compared in (3.86) and (3.67). As a result of such a comparison we obtain the initial moments of uniform distribution

$$m_k = \frac{1}{k+1}\cdot\frac{b^{k+1} - a^{k+1}}{b-a}. \tag{3.87}$$

From (3.87), as special cases, we find the mean value and the dispersion of a random variable uniformly distributed over the interval (a, b)

$$m_1 = \frac{b+a}{2},\ M_2 = m_2 - m_1^2 = \frac{(b-a)^2}{12}. \tag{3.88}$$

If $a = -b$, then

$$m_{2k+1} = M_{2k+1} = 0,\ m_{2k} = M_{2k} = \frac{b^{2k}}{2k+1}. \tag{3.89}$$

Let us finally examine the characteristic function of the values of a random-phase sinusoid. With (3.16) in mind, we find

$$\Theta(v) = \frac{1}{\pi a}\int_{-a}^{a}\frac{e^{ivx}}{\sqrt{1-\left(\frac{x}{a}\right)^2}}dx = \frac{1}{\pi}\int_{-1}^{+1}\frac{e^{iauv}}{\sqrt{1-u^2}}du = J_0(av), \tag{3.90}$$

where $J_0(av)$ is a zero-order Bessel function.

The corresponding moments of distribution are equal to

$$m_k = M_k = \frac{1}{i^k}\left[\frac{d^k J_0(av)}{dv^k}\right]_{v=0}. \tag{3.91}$$

By successive differentiation we find

$$m_1 = 0,\ M_2 = \frac{a^2}{2},\quad M_3 = 0,\quad M_4 = \frac{3a^4}{8}. \tag{3.92}$$

Let us note that formula (3.90) follows directly from the definition of a characteristic function and from (3.73) (cf. also pp. 90 - 91).

$$\Theta(v) = m_1\{e^{iva\sin\omega\xi}\} = \frac{\omega}{2\pi}\int_{-\frac{\pi}{\omega}}^{\frac{\pi}{\omega}} e^{iva\sin\omega x}dx =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iav\sin z}dz = J_0(av).$$

9. Distribution Function of Sum of Normally Distributed Random Variables

Let us examine the sum of n mutually independent variables, each of which is normally distributed with the parameters a_k and σ_k (k = 1, 2,...n).

The characteristic function of this sum, according to (3.71) and (3.79) is equal to

$$\Theta(v) = \prod_{k=1}^{n}\Theta_{1k}(v) = \prod_{k=1}^{n} e^{ia_kv - \frac{\sigma_k^2 v^2}{2}}$$

or

$$\Theta(v) = e^{iv\sum_{k=1}^{n}a_k - \frac{v^2}{2}\sum_{k=1}^{n}\sigma_k^2}$$

Let us designate

$$a = \sum_{k=1}^{n} a_k,\quad \sigma^2 = \sum_{k=1}^{n}\sigma_k^2. \tag{3.93}$$

Then

$$\Theta(v) = e^{iav - \frac{\sigma^2 v^2}{2}}. \tag{3.94}$$

Comparing (3.79) and (3.94), we conclude that the characteristic function of a sum of normally distributed random variables coincides with the characteristic function of a normally distributed random variable, but with the mean value and dispersion determined by formula (3.93).

Thus the sum of any number of independent, normally distributed random variables is also distributed normally, the mean value of this sum being equal to the sum of the mean values of the terms, and the dispersion of the sum being equal to the sum of the dispersions of the terms. These properties of mean values and dispersions are special cases of (3.52) and (3.58).

Let us apply the method of characteristic functions to the determination of the distribution function of the sum $\xi_1 + \xi_2$ of two dependent, random variables, distributed according to the normal law. For this, employing (3.74) and (2.51), we first compute the two-dimensional characteristic function of these random variables:

$$\Theta_2(v_1, v_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times$$
$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)}\left\{\frac{(x_1-a_1)^2}{\sigma_1^2} - 2r\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right\}} \times$$
$$\times e^{i(v_1x_1+v_2x_2)}\,dx_1dx_2.$$

As a result of computing the integral (cf. Appendix V), we obtain

$$\Theta_2(v_1, v_2) = e^{i(a_1v_1+a_2v_2) - \frac{1}{2}(\sigma_1^2v_1^2 + 2r\sigma_1\sigma_2v_1v_2 + \sigma_2^2v_2^2)} \tag{3.95}$$

If in (3.95) it is assumed that $v_2 = 0$, then in accordance with (3.74') we obtain the one-dimensional characteristic function (3.79) of a normally distributed random variable.

To determine the characteristic distribution function $\Theta(v)$ of the sum of two random variables, distributed according to the normal law, it is sufficient now to make use of formula (3.76):

$$\Theta(v) = \Theta_2(v, v) = e^{iv(a_1+a_2) - \frac{v^2}{2}(\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2)}. \tag{3.96}$$

It can be seen from a comparison of (3.96) with (3.79) that the obtained function Θ ($\underline{v}$) is a characteristic function, corresponding to the normal law of distribution, with a mean value of $a_1 + a_2$ and a dispersion of $\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2$.

Thus the sum of two dependent, normally distributed random variables is also normally distributed, the mean value of this sum being equal to the sum of the mean values of the terms, and the dispersion σ^2 of the sum being equal to

$$\sigma^2 = \sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2.$$

The obtained result is generalized for the sum of an arbitrary number of dependent random variables, linked by an $\underline{n}$-dimensional normal distribution (2.48). It can be shown that an $\underline{n}$-dimensional characteristic function, which corresponds to this distribution, has the form of

$$\Theta_n(v_1, v_2, \ldots v_n) = e^{i\sum_{k=1}^{n} a_k v_k - \frac{1}{2}\sum_{l=1}^{n}\sum_{k=1}^{n} \sigma_l \sigma_k r_{lk} v_l v_k} \tag{3.95'}$$

where $\underline{r_{lk}}$ is the coefficient of correlation between the random variables ξ_l and ξ_k ($r_{lk} = r_{kl}$, $r_{kk} = 1$). Then, employing (3.76), we find the characteristic function of the sum of $\underline{n}$ normally distributed random variables

$$\Theta(v) = \Theta_n(v, v, \ldots v) = e^{iv\sum_{k=1}^{n} a_k - \frac{v^2}{2}\sum_{l=1}^{n}\sum_{k=1}^{n} \sigma_l \sigma_k r_{lk}} \tag{3.96'}$$

From (3.96'), it follows that the sum of the indicated random variables is also distributed normally, the mean value of the sum being equal to the sum of the mean values of the terms, while the dispersion σ^2 of the sum is equal to

$$\sigma^2 = \sum_{l=1}^{n}\sum_{k=1}^{n} \sigma_l \sigma_k r_{lk}.$$

Let us also find the distribution function of the summation $\chi^2 = \frac{1}{\sigma^2}\sum_{K=1}^{n}(\xi_K - a)^2$, where $\xi_1, \xi_2, \ldots \xi_n$ are independent, normally distributed random variables with mean values of $\underline{a}$ and dispersion σ^2. The characteristic function of each term of this summation, i.e., of the square of the normalized deviation ξ_k, is determined

by formula (3.81), if it is assumed therein that $\sigma = 1$. Therefore the characteristic function of χ^2 is equal to

$$\Theta_{1n}(v) = (1 - 2iv)^{-\frac{n}{2}}, \tag{3.97}$$

wherefrom by means of an inverse Fourier transformation we calculate the desired distribution function of the summation:

$$w_{1n}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ivx}}{(1-2iv)^{n/2}} dv =$$
$$= \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geqslant 0, \\ 0, & x < 0. \end{cases} \tag{3.98}$$

Distribution (3.98) is known as the χ^2 distribution. Curves of this distribution for several values of n are presented in Fig. 3.26. When $n \leqslant 2$, function (3.98) decreases monotonously for positive values of the argument, and when $n > 2$ it has a single maximum at the point of $x = n - 2$. With an increase in the number of n, this distribution curve approaches the curve of normal distribution.

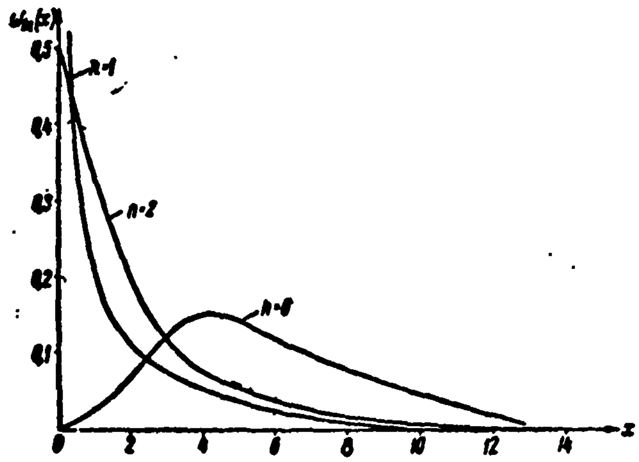


Fig. 26. The χ^2 distribution

A special case of (3.98) for n = 1 is the distribution function (3.14), examined in Section 2.

10. Distribution Function of Sum of Uniformly Distributed Random Variables.

Let us find the distribution function of two independent random variables, each of which is uniformly distributed over the interval of (a, b). The characteristic function $\Theta_{12}(v)$ of this sum is, in accordance with (3.71), equal to the square of characteristic function (3.85), i.e., to

$$\Theta_{12}(v) = -\frac{(e^{ibv} - e^{iav})^2}{v^2 (b-a)^2}. \tag{3.99}$$

From (3.99) we calculate, by means of an inverse Fourier transformation, the desired distribution function of a sum of random variables:

$$w_{12}(x) = \frac{1}{2\pi} \cdot \frac{-1}{(b-a)^2} \int_{-\infty}^{\infty} \frac{(e^{ibv} - e^{iav})^2}{v^2} e^{-ixv} dv. \tag{3.100}$$

By substituting the variable $x = y + a + b$ we reduce (3.100) to

$$w_{12}(y) = \frac{2}{\pi (b-a)} \int_0^{\infty} \cos \frac{2yu}{b-a} \frac{\sin^2 u}{u^2} du. \tag{3.101}$$

The integral obtained is tabular and is equal to

$$w_{12}(y) = \begin{cases} \frac{1}{b-a}\left(1 - \frac{y}{b-a}\right), & 0 < y < b-a, \\ 0, & y > b-a. \end{cases}$$

It may be seen from (3.101) that the function $\omega_{12}(y)$ is even, therefore with a negative argument its values are calculated from the equality

$$w_{12}(-y) = w_{12}(y).$$

Returning from the variable y to the variable x, we obtain the following expression for the desired distribution function of a sum of two independent, uniformly distributed random variables

$$
w_{12}(x) = \begin{cases} 0, & x \leqslant 2a, \\ \dfrac{x-2a}{(b-a)^2}, & 2a \leqslant x \leqslant a+b, \\ \dfrac{2b-x}{(b-a)^2}, & a+b \leqslant x \leqslant 2b, \\ 0, & x \geqslant 2b. \end{cases} \tag{3.102}
$$

This function takes the form of a triangle, and this distribution is therefore called a <u>triangular</u> distribution (sometimes, a <u>Simpson</u> distribution).

It is possible analogously to find the distribution function of three independent random variables, each of which is uniformly distributed over the interval of ($\underline{a}$,$\underline{b}$). The characteristic function of this sum is equal to the cube of characteristic function (3.85), i.e., to

$$
\Theta_{13}(v) = -\frac{\left(e^{ibv} - e^{iav}\right)^3}{iv^3 (b-a)^3}. \tag{3.103}
$$

Subjecting (3.103) to an inverse Fourier transformation, similar to the preceding one, we find the desired distribution function of the sum of random variables

$$
w_{13}(x) = \begin{cases} 0, & x \leqslant 3a, \\ \dfrac{1}{2}\dfrac{(x-3a)^2}{(b-a)^3}, & 3a \leqslant x \leqslant 2a+b, \\ \dfrac{1}{2}\dfrac{(x-3a)^2 - 3[x-(b+2a)]^2}{(b-a)^3}, & 2a+b \leqslant x \leqslant a+2b, \\ \dfrac{1}{2}\dfrac{(3b-x)^2}{(b-a)^3}, & a+2b \leqslant x \leqslant 3b, \\ 0, & x \geqslant 3b. \end{cases} \tag{3.104}
$$

Fig. 27 shows an initial uniform distribution, and also the distributions of a sum of two and a sum of three independent, uniformly distributed random variables.

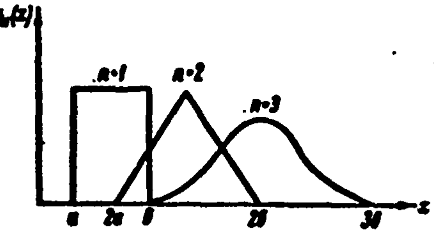


Fig. 27. Distribution of a sum of independent, uniformly distributed random variables.

With an increase in the number of n terms, the distribution function rapidly approaches a normal distribution function with the parameters of $a_n = \frac{n(a+b)}{2}$, $\sigma_n = (b-a)\sqrt{\frac{n}{12}}$ [compare (3.88)].

11. Distribution Function of Randomly Phased Sum of Harmonic Vibrations

Let us find the distribution function of a sum of independent harmonic vibrations with constant amplitudes and with random, uniformly distributed phases:

$$\xi = a_1 \cos\omega_1\xi_1 + a_2 \cos\omega_2\xi_2 + \ldots + a_n \cos\omega_n\xi_n. \tag{3.105}$$

The distribution function of each of the terms in sum (3.105) was determined in Section 2, and the characteristic function corresponding to it is provided by formula (3.90). This function depends only on the amplitude a_k, but does not depend on the frequency ω_k. In accordance with (3.71) the characteristic function $\Theta_{1n}(v)$ of the sum ξ is equal to the product of the characteristic functions of the terms and has the form of

$$\Theta_{1n}(v) = \prod_{k=1}^{n} J_0(a_k v). \tag{3.106}$$

From (3.106) we calculate, by means of an inverse Fourier transformation, the distribution function $\omega_{1n}(x)$ of the randomly phased sum of independent harmonic vibrations:

$$\begin{gathered} w_{1n}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\prod_{k=1}^{n} J_0(a_k v)\, e^{-ivx}\, dv,\ |x| < \sum_{k=1}^{n} a_k = A, \\ w_{1n}(x) = 0,\ |x| > A. \end{gathered} \tag{3.106'}$$

Since $\omega_{1n}(x) \equiv 0$ when $|x| > A$, this function may, over the interval of $|x| < A$, be expanded into a Fourier series:

$$w_{1n}(x) = \sum_{r=-\infty}^{\infty} c_r e^{\frac{i\pi r x}{A}}.$$

where

$$c_r = \frac{1}{2A}\int_{-A}^{A} w_{1n}(x)\, e^{-\frac{i r \pi x}{A}}\, dx = \frac{1}{2A}\prod_{k=1}^{n} J_0\left(\frac{a_k \pi r}{A}\right).$$

Thus

$$w_{1n}(x) = \frac{1}{2A}\sum_{r=-\infty}^{\infty} e^{\frac{i r \pi x}{A}} \prod_{k=1}^{n} J_0\left(\frac{a_k \pi r}{A}\right) =$$

$$= \frac{1}{2A}\left[1 + 2\sum_{r=1}^{\infty} \cos\frac{\pi r x}{A}\cdot\prod_{k=1}^{n} J_0\left(\frac{a_k \pi r}{A}\right)\right],\ |x| < A. \qquad (3.107)$$

If the amplitudes of all the harmonic vibrations are the same ($a_k = a$, $A = na$), then

$$w_{1n}(x) = \frac{1}{2na}\left[1 + 2\sum_{r=1}^{\infty} \cos\frac{\pi r x}{na} J_0^n\left(\frac{r\pi}{n}\right)\right],\ |x| < na, \qquad (3.107')$$

The probability, that the absolute value of the sum of the independent vibrations will not exceed λ times the amplitude of one term, would be

$$P\{|\xi| < \lambda a\} = 2\int_0^{\lambda a} w_{1n}(x)\, dx = \frac{\lambda}{n} + \frac{2}{na}\sum_{r=1}^{\infty} J_0^n\left(\frac{r\pi}{n}\right)\int_0^{\lambda a} \cos\frac{\pi r x}{na}\, dx,$$

or

$$P\{|\xi| < \lambda a\} = \frac{\lambda}{n} + \frac{2}{\pi}\sum_{r=1}^{\infty} \frac{1}{r} J_0^n\left(\frac{r\pi}{n}\right)\sin\frac{\pi r \lambda}{n},\ \lambda < n. \qquad (3.108)$$

It can be shown that, with an increase in the number of n terms, the distribution (3.107) approaches the normal. From (2.18) it follows that, for a normal random variable η with a dispersion of $\frac{na^2}{2}$, the probability is

$$P\{|\eta| < \lambda a\} = 2F\left(\lambda\sqrt{\frac{2}{n}}\right) - 1. \qquad (3.108')$$

Table 4 shows the values of probabilities computed according to formulas (3.108) and (3.108') for n = 10; it is evident from this that the randomly-phased sum of 10 harmonic vibrations of the same amplitude is sufficiently close in its statistical

characteristics, to a normal random variable.

Table 4

	λ	1	2	3	4	5	6	7
p	according to (3.108)	0,3393	0,6221	0,8168	0,9267	0,9766	0,9944	0,9991
	according to (3.108')	0,3473	0,6319	0,8130	0,9281	0,9756	0,9931	0,9985

The distribution function of the randomly phased sum of two harmonic vibrations of equal amplitude is expressed as an elliptical integral. From (3.106), when $n = 2$ we have $\Theta_{12}(v) = J_0^2(av)$.

Presenting the square of the Bessel function in the form* of

$$J_0^2(av) = \frac{2}{\pi}\int_0^{\frac{\pi}{2}} J_0(2av\sin\psi)\,d\psi$$

and employing (3.106'), we obtain, changing the order of integration with respect to v and ψ,

$$w_{12}(x) = \frac{2}{\pi^2}\int_0^{\frac{\pi}{2}}\int_0^{\infty} J_0(2av\sin\psi)\cos vx\,dv\,d\psi.$$

The inner integral is a Fourier transformation of a Bessel function, therefore, considering (3.90), we find that

$$\int_0^{\infty} J_0(2av\sin\psi)\cos vx\,dv = \begin{cases} \dfrac{1}{\sqrt{4a^2\sin^2\psi - x^2}}, & |x| < 2a\sin\psi, \\ 0, & |x| > 2a\sin\psi. \end{cases}$$

Thus,

$$w_{12}(x) = \frac{2}{\pi^2}\int_{\arcsin\frac{|x|}{2a}}^{\frac{\pi}{2}} \frac{d\psi}{\sqrt{4a^2\sin^2\psi - x^2}}, \quad |x| < 2a,$$

$$w_{12}(x) = 0, \quad |x| > 2a.$$

Substituting, finally, the variable of integration $u = \sin\psi$, we obtain the desired distribution function of the randomly phased sum of two independent harmonic vi-

* cf. G. N. Vatson. Teoriya besselevykh funktsiy (Theory of Bessel Functions), Moskva, For. Lit. Pub. Hse, 1949, (i.e., G. N. Watson, "A Treatise on the Theory of Bessel Functions", 2nd Ed., N. Y., 1944).

brations in the form* of

$$w_{12}(x)=\frac{2}{\pi^2}\int\limits_{\frac{|x|}{2a}}^{1}\frac{du}{2a\sqrt{(1-u^2)\left(u^2-\frac{x^2}{4a^2}\right)}}=\frac{1}{a\pi^2}K\left(\sqrt{1-\frac{x^2}{4a^2}}\right)$$

when $|x|<2a$,

$$w_{12}(x)=0 \text{ when } |x|>2a, \tag{3.109}$$

where k is a complete elliptical integral of the first kind.

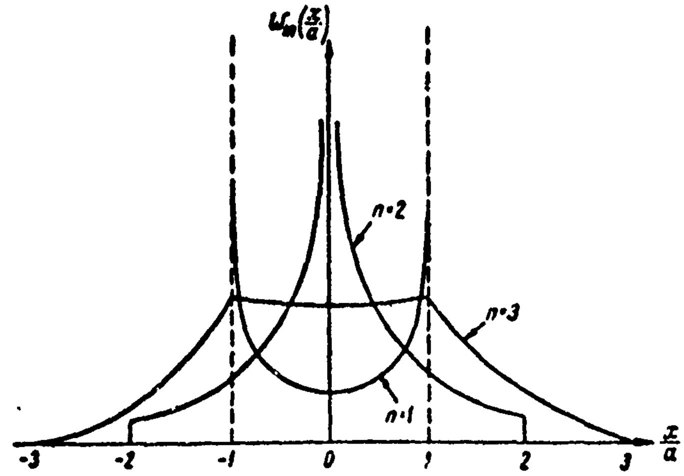


Fig. 28. Distribution of a randomly phased sum of independent vibrations.

Fig. 28 shows the initial randomly phased distribution of harmonic vibration, and also the sums of two and three independent vibrations. With an increase in the number of n terms, the distribution function of a sum rapidly approaches the normal.

12. Generating Function of Discrete Random Variable

For a discrete random variable which takes integer values of r with probabilities of P_r, a role analogous to that of the characteristic function of a continuous random variable is played by the generating function, defined by the equality

$$\Theta(v)=\sum_{r=0}^{\infty}v^r p_r. \tag{3.110}$$

By means of the generating function it is easy to calculate moments of distribution, as well as distributions of sums of discrete random variables. By successive dif-

*Let us note that distribution $\omega_{12}(x)$ in the form of integral (3.109) is obtained directly from (3.24) as a convolution of the two distribution functions (3.16).

ferentiation we obtain

$$\Theta'(v)=\sum_{r=0}^{\infty} r v^{r-1} p_r,$$

$$\Theta''(v)=\sum_{r=0}^{\infty} r(r-1) v^{r-2} p_r,$$

wherefrom

$$\Theta'(1)=\sum_{r=0}^{\infty} r p_r = m_1, \tag{3.111}$$

$$\Theta''(1)=\sum_{r=0}^{\infty} r(r-1) p_r = m_2 - m_1 \text{ etc.} \tag{3.112}$$

It is possible to show that the generating function of a sum of independent random variables is equal to the product of the generating functions of the terms.

For a discrete random variable distributed according to Poisson's law the generating function is

$$\Theta(v)=\sum_{r=0}^{\infty} v^r \frac{\lambda^r}{r!} e^{-\lambda} = e^{\lambda(v-1)}. \tag{3.113}$$

The generating function of a sum of two random variables, distributed according to Poisson's law with parameters of λ_1 and λ_2 is equal to

$$\Theta(v)=\Theta_1(v)\Theta_2(v)=e^{\lambda_1(v-1)}e^{\lambda_2(v-1)}=e^{(\lambda_1+\lambda_2)(v-1)}, \tag{3.114}$$

i.e., is equal to the generating function of a Poisson distribution with a parameter of $\lambda_1 + \lambda_2$.

Thus the Poisson distribution, just as the normal, reproduces itself in the addition of independent random variables.

From (3.113) (3.111) and (3.112) it is not difficult to find the mean value and dispersion of the Poisson distribution

$$m_1=\Theta'(1)=\lambda,$$
$$M_2=m_2-m_1^2=\Theta''(1)+m_1-m_1^2=\lambda^2+\lambda-\lambda^2=\lambda.$$

LITERATURE

1. B. V. Gnedenko. Kurs teorii veroyatnostey (Course in Probability Theory), Moskva, Gostekhizdat, 1954.
2. G. Kramer (i.e., Harold Cramer), Matematicheskiye metody statisiki (Mathematical Methods of Statistics), Moskva, For. Lit. Pub. Hse., 1948.
3. Dunin-Barkovskiy, I. V. and Smirnov, N. V. Teoriya veroyatnostey i matematicheskaya statistika v tekhnike (Probability Theory and Mathematical Statistics in Engineering), Moskva, Gostekhizdat, 1955.
4. M. Slack. The Probability Distributions of Sinusoidal Oscillations Combined in Random Phase. Journ. IEE, p. III, 93, No. 22. March 1946.
5. W. R. Bennett. Distribution of the Sum of Randomly Phased Components. Quart. Appl. Math. V, No. 4, January 1948.
6. A. Khal'd. Matematicheskaya statistika s tekhnicheskimi prilozheniyami (Mathematical Statistics with Technical Applications). Moskva, For. Lit. Pub. Hse., 1956.
7. S. O. Rice. Mathematical Analysis of Random Noise, B ST J, 23, No. 3. July 1944; No. 1, Jan. 1945 (a /Russian/ translation is avaliable in the collection "Teoriya peredachi elektricheskikh signalov pri nalichii pomekh" (Theory of the Transmission of Electrical Signals in the Presence of Interference), Moskva, For. Lit. Pub. Hse., 1953).

Chapter IV

LIMITING THEOREMS

1. The Lyapunov Theorem

One of the important results of Section 9 of the preceding chapter was the following: the sum of independent, normally distributed random variables is also normally distributed. On the other hand, the distribution curve of a sum of three uniformly distributed random variables, shown in Fig. 27, approaches the normal distribution curve.

The same phenomenon has also been noted with respect to the distribution of a randomly phased sum of harmonic vibrations, as well as to the χ^2 distribution (Figs. 26 and 28).

Further, in a succession of independent tests, the number of k occurrences of an event in n tests may be regarded as the summation of $\sum_{r=1}^{n} \xi_r$ independent discrete random variables, each of which can take only two values: unity with a probability of p, and zero with a probability of $q = 1 - p$. According to Laplace's asymptotic formula given a large values of n the distribution of the number of occurrences of events, i.e., the distribution of a sum of independent, discrete random variables of an indicated type, may with some accuracy be considered as normal.

The question arises: cannot this property of a sum of individual random variables, proved in several special cases, be generalized to become an arbitrary law of distribution of the terms. The answer to this is affirmative, and constitutes the contents of the so-called central limiting theorem of probability theory, proved first by Lyapunov and subsequently refined by other mathematicians. The great importance which this theorem has, in natural science and in engineering (including radio engineering), is linked with the fact that in practice it is very often necessary to investigate phenomena which are affected by a large number of independently acting random factors.

As has been pointed out in the preceding chapter, the most convenient and simple method of studying a sum of independent random variables is the method of characteristic functions. This method will be used below for proving the Lyapunov theorem with respect to sums of independent, continuous random variables. It may also be employed for discrete random variables.

Let us then examine the summation of $\xi = \sum_{r=1}^{n} \xi_r$ independent random variables, the distribution functions of whose individual terms are the same and equal to $\omega(x)$. Then the mean value, a, and the dispersion, σ^2, of each term, are equal to

$$a = \int_{-\infty}^{\infty} x w(x)\,dx, \quad \sigma^2 = \int_{-\infty}^{\infty} (x-a)^2 w(x)\,dx,$$

and the mean value and dispersion of the summation, ξ, are, in accordance with (3.52) and (3.58) equal to

$$m_1\{\xi\} = na, \quad M_2\{\xi\} = n\sigma^2. \tag{4.1}$$

We shall be seeking the distribution function of the normalized deviation of the sum, i.e., of the random variable

$$\eta = \frac{\xi - na}{\sigma\sqrt{n}} = \sum_{r=1}^{n} \frac{\xi_r - a}{\sigma\sqrt{n}}. \tag{4.2}$$

For this it is useful first to find the characteristic function of each term of the sum (4.2), and then to employ the rule already known to us, namely that the characteristic function of a sum of independent random variables is equal to the product of the characteristic functions of the terms.

Since for any term, ξ_r, the distribution function is the same and equal to $\omega(x)$, the characteristic function of the deviations $\xi_r - a$ ($r = 1, 2, \ldots n$) will be the same and equal to

$$\Theta(v) = \int_{-\infty}^{\infty} w(x) e^{iv(x-a)}\,dx. \tag{4.3}$$

Expanding $e^{iv(x-a)}$ into an exponential series, and changing the order of summation and integration, we find

$$\Theta(v)=\int_{-\infty}^{\infty} w(x)\sum_{k=0}^{\infty}\frac{i^k v^k (x-a)^k}{k!}dx=$$

$$=\sum_{k=0}^{\infty}\frac{i^k v^k}{k!}\int_{-\infty}^{\infty}(x-a)^k w(x)\,dx,$$

or

$$\Theta(v)=\sum_{k=0}^{\infty}\frac{i^k M_k}{k!}v^k, \tag{4.4}$$

where M_k are central moments of the distribution $\omega(x)$. Expansion (4.4) is obtained from (3.67) through a replacement of m_k by M_k, since here $\Theta(v)$ is the characteristic function of the normalized deviation of a random variable.) The characteristic function of the normalized deviation $\frac{\xi_k - a}{\sigma\sqrt{n}}$ is, according to (3.64), equal to

$$\Theta\left(\frac{v}{\sigma\sqrt{n}}\right)=\sum_{k=0}^{\infty}\frac{i^k M_k}{\sigma^k n^{k/2} k!}v^k. \tag{4.5}$$

The desired characteristic function of the sum η is, according to (3.71'), equal to

$$\Theta_{1n}(v)=\left[\Theta\left(\frac{v}{\sigma\sqrt{n}}\right)\right]^n. \tag{4.6}$$

Substituting into (4.6) the expression for $\Theta\left(\frac{v}{\sigma\sqrt{n}}\right)$ from (4.5), we obtain ($M_1 = 0$, $M_2 = \sigma^2$)

$$\Theta_{1n}(v)=\left(\sum_{k=0}^{\infty}\frac{i^k M_k v^k}{\sigma^k n^{k/2} k!}\right)^n=\left(1-\frac{v^2}{2n}-\frac{iM_3 v^3}{6\sigma^3 n^{3/2}}+\frac{M_4 v^4}{24\sigma^4 n^2}+\dots\right)^n. \tag{4.7}$$

The approach to the limit of $n \to \infty$ in (4.7), results in an indeterminate form of the type 1^{∞}. To clear up this indeterminacy, we examine $\ln \Theta_{1n}(v)$. Then from (4.7) it follows that

$$\ln\Theta_{1n}(v)=n\ln\left(1-\frac{v^2}{2n}-\frac{iM_3 v^3}{6\sigma^3 n^{3/2}}+\frac{M_4 v^4}{24\sigma^4 n^2}+\dots\right). \tag{4.8}$$

Expanding the logarithm into a series, we obtain

$$\ln \Theta_{1n}(v) = -\frac{v^2}{2} - \frac{iM_3 v^3}{6\sigma^3 \sqrt{n}} + \frac{M_4 v^4}{24\sigma^4 n} - \frac{v^4}{8n} + \ldots \tag{4.9}$$

Approaching the limit when $n \longrightarrow \infty$, we find from (4.9) that

$$\lim_{n\to\infty} \ln \Theta_{1n}(v) = -\frac{v^2}{2},$$

wherefrom

$$\lim_{n\to\infty} \Theta_{1n}(v) = e^{-\frac{v^2}{2}}. \tag{4.10}$$

From (4.10) it follows that the distribution function $\omega_{1n}(x)$ for the normalized deviation of a sum of independent random variables tends toward the normal as the number of terms is increased

$$\lim_{n\to\infty} w_{1n}(x) = w(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{4.11}$$

or

$$w_{1n}(x) \sim \frac{}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{4.12}$$

Employing the expression (4.1) of the mean value and dispersion of a sum, it is not difficult to pass from the distribution function $\omega_{1n}(x)$, of a normalized deviation, to the distribution function $W_{1n}(x)$ of a sum of independent random variables

$$W_{1n}(x) \sim \frac{1}{\sqrt{2\pi n}\,\sigma} e^{-\frac{(x-na)^2}{2n\sigma^2}}. \tag{4.13}$$

In this manner it has been proven that the law of distribution of a sum of independent random variables having the same distribution function tends, as the sum of the terms is increased, toward the normal, regardless of what the distribution of the terms had been.

With certain supplementary assumptions, the Lyapunov theory is valid also in the case when the distribution functions of the terms are unequal. It may be expanded even to certain cases when the terms ξ_K are not independent [5].

On the other hand, the distribution of a sum of independent random variables

does not always converge towards the normal. In the derivation presented above it was assumed that the mean values and dispersions of the terms exist and are finite. However, this condition might not be met even in practice. Suppose, for instance, there is studied the distribution of a sum of random variables, each of which is equal to $\eta_r = \frac{1}{\xi_r^2}$, with ξ_r being distributed according to the Rayleigh law (2.61)*. Then the distribution function of the random variable η_r is, in accordance with (3.7), equal to

$$W(y) = \frac{1}{\sigma^2 y^2} e^{-\frac{1}{2\sigma^2 y}}.$$

A mean value and a distribution for each term (i.e., for the random variable η_r) do not exist, since the integrals

$$m_1 = \frac{1}{\sigma^2}\int_0^\infty \frac{1}{y} e^{-\frac{1}{2\sigma^2 y}} dy, \quad m_2 = \frac{1}{\sigma^2}\int_0^\infty e^{-\frac{1}{2\sigma^2 y}} dy$$

diverge.

The general conditions (necessary and sufficient), for the convergence of the distribution function of sums of independent random variables towards the normal law, have been found in recent years [5]. Parallel to this there arose the question as t.. ..at law, besides the normal, can be limiting for sums of independent random variables. It turned out that the limiting laws are far from exhausted by the normal law. A large class of normalized sums of independent random variables (including those for which terminal numerical characteristics do not exist) have distributions tending toward the so-called stable laws. Each stable law may be reduced to a canonical form, determined by the characteristic function

$$\Theta(u) = e^{-\left[|u|^\alpha\left(1 + i\beta\frac{u}{|u|}\operatorname{tg}\frac{\pi\alpha}{2}\right)\right]}, \quad \alpha \neq 1,$$

$$\Theta(u) = e^{-\left[|u|\left(1 + \frac{2i\beta}{\pi}\frac{u}{|u|}\ln|u|\right)\right]}, \quad \alpha = 1.$$

* Sums of this type occur in research on the influence of propagation conditions (fading) in radio-relay communication systems (cf, also Section 5 of this chapter).

When $\alpha = 2$, the stable law corresponds to the normal law.

2. Estimate of the Rate of Convergence Towards the Normal Law.

For practical applications of the Lyapunov theorem it is necessary to estimate the asymptotic equalities (4.12) and (4.13) in relation to the number of n terms and to the form of the distribution function $\omega(x)$.

A correction to the normal law for the distribution function of a sum of independent random variables is obtained from an examination of expression (4.8). The function $\ln\theta_{1n}(v)$ is an exponential series in terms of v, the coefficients of which depend on n and on the central moments of the initial distribution of the terms. Depending on the precision required for the evaluation of the approach of $\omega_{1n}(x)$ to the normal law, it is possible to restrict one's self to the necessary number of the terms in this series. If, for example, it is required to retain terms of an order not exceeding $\frac{1}{n}$, then from (4.9) we find

$$\Theta_{1n}(v) = e^{-\frac{v^2}{2}} e^{-\frac{iM_3 v^3}{6\sigma^3\sqrt{n}} + \frac{\left(\frac{M_4}{\sigma^4}-3\right)v^4}{24n}}.$$

Introducing the coefficients of asymmetry and excess of the initial distribution $n(x)$ according to formulas (2.82) and (2.84) and expanding the exponential function into a series, we obtain with a precision of the order of $O(n^{-3/2})$

$$\Theta_{1n}(v) = e^{-\frac{v^2}{2}}\left(1 - \frac{ik}{6\sqrt{n}}v^3 + \frac{\gamma}{24n}v^4 - \frac{k^2}{2\cdot 36n}v^6\right). \tag{4.14}$$

From (4.14) we find with the same precision the distribution function of the normalized deviation (4.2)

$$w_{1n}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Theta_{1n}(v)\, e^{-ivx}\, dv =$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(1 - \frac{ik}{6\sqrt{n}}v^3 + \frac{\gamma}{24n}v^4 - \frac{k^2}{72n}v^6\right) e^{-ivx - \frac{v^2}{2}}\, dv. \tag{4.15}$$

In the right side of (4.15) there stands the sum of four integrals, the first

of which, according to (3.79), is equal to

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ivx-\frac{v^2}{2}}\,dv=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}. \qquad (4.16)$$

The remaining two are obtained from (4.16) by differentiating with respect to x:

$$\frac{d^3}{dx^3}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ivx-\frac{v^2}{2}}\,dv\right)=\frac{(-i)^3}{2\pi}\int_{-\infty}^{\infty} v^3 e^{-ivx-\frac{v^2}{2}}\,dv=$$
$$=\frac{1}{\sqrt{2\pi}}\frac{d^3}{dx^3}\left(e^{-\frac{x^2}{2}}\right)=-H_3(x)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \qquad (4.17)$$

$$\frac{d^4}{dx^4}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ivx-\frac{v^2}{2}}\,dv\right)=\frac{(-i)^4}{2\pi}\int_{-\infty}^{\infty} v^4 e^{-ivx-\frac{v^2}{2}}\,dv=$$
$$=\frac{1}{\sqrt{2\pi}}\frac{d^4}{dx^4}\left(e^{-\frac{x^2}{2}}\right)=H_4(x)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \qquad (4.18)$$

$$\frac{d^6}{dx^6}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ivx-\frac{v^2}{2}}\,dv\right)=\frac{(-i)^6}{2\pi}\int_{-\infty}^{\infty} v^6 e^{-ivx-\frac{v^2}{2}}\,dv=$$
$$=\frac{1}{\sqrt{2\pi}}\frac{d^6}{dx^6}\left(e^{-\frac{x^2}{2}}\right)=H_6(x)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \qquad (4.18')$$

where $H_3(x)$, $H_4(x)$, $H_6(x)$ are Hermite polynomials (cf. Appendix VII).

Substituting (4.16), (4.17), (4.18) and (4.18') into (4.15), we find

$$w_{1n}(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\left[1-\frac{k}{6\sqrt{n}}H_3(x)+\frac{\gamma}{24n}H_4(x)+\frac{k^2}{72n}H_6(x)\right]. \qquad (4.19)$$

From (4.19) it follows that the distribution of the normalized deviation of a sum of random variables, which have symmetrical distribution functions of ($k = 0$), converges upon the normal distribution more rapidly than the normalized deviations of a sum of random variables for which the distribution functions are asymmetrical ($k \neq 0$). From (4.19) we find also the expression of the distribution function of the sum of independent random variables itself

$$W_{1n}(x)=\frac{1}{\sqrt{2\pi n}\,\varepsilon}e^{-\frac{(x-na)^2}{2n\varepsilon^2}}\left[1-\frac{k}{6\sqrt{n}}H_3\left(\frac{x-na}{\varepsilon\sqrt{n}}\right)+\right.$$

$$\left.+\frac{\gamma}{24n}H_4\left(\frac{x-na}{\varepsilon\sqrt{n}}\right)+\frac{k^2}{72n}H_6\left(\frac{x-na}{\varepsilon\sqrt{n}}\right)\right]. \tag{4.20}$$

3. Examples.

As examples illustrating the application of the formula obtained above, let us find and compare the distribution functions of the normalized deviation of a sum of seven random variables for three special cases: a) the terms are distributed according to the Rayleigh law, b) the terms are distributed uniformly, with a zero mean value, c) the terms are distributed according to the law of (3.16) (i.e., a sum of sinusoids with random phases is examined). We assume that the dispersions of the terms are the same in all three cases.

Table 5 shows numerical characteristics (mean values, dispersion, coefficients of asymmetry and excess) of random variables for the indicated distributions, as well as for normal distribution.

Employing (4.19), we find the distribution functions of the normalized deviations of the sums of random variables in the indicated special cases

$$w_a(x)=\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}(0{,}995-0{,}123x+0{,}0106x^2+0{,}041x^3-0{,}0018x^4)$$

(Rayleigh distribution of the terms),

$$w_b(x)=\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}(0{,}979+0{,}043x^2-0{,}007x^4)$$

(uniform distribution of the terms),

$$w_c(x)=\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}(0{,}973+0{,}054x^2-0{,}009x^4)$$

(sum of sinusoids with random phases).

Table 5

Numerical Characteristic	Distribution			
	Normal	Rayleigh	Uniform	Sinusoid
Mean value	0	$\sqrt{\frac{2\pi}{4-\pi}}$	0	0
Dispersion	2	2	2	2
Coefficient of Asymmetry	0	0,63	0	0
Coefficient of excess	0	−0,3	−1,2	−1,5

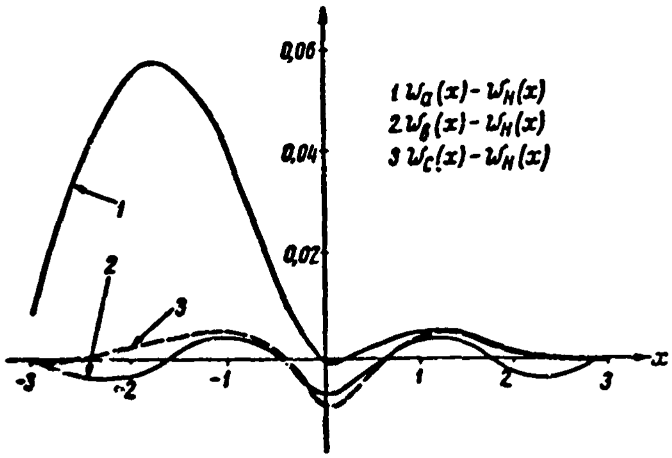


Fig. 29. Convergence of sums of independent random variables to the normal. The figure shows the differences between distribution functions of the normalized deviations of

1) the sum of seven Rayleigh random variables and the normal,

2) the sum of seven uniformly distributed random variables and the normal,

3) the sum of seven vibrations with random phases and the normal.

In Fig. 29 is plotted the differences between the distribution functions $\omega_a(x)$, $\omega_b(x)$, $\omega_c(x)$ and the limiting function $\omega_n(x)$ of normal distribution. From the figure it is evident that the distributions of a sum of sinusoids with a random phase,

where a and b are the mean values, while σ_1^2 and σ_2^2 are the dispersions of each of the random variables ξ_r and η_r respectively.

The two-dimensional characteristic function of either of the vectors $(\xi_r - a; \eta_r - b)$ is equal to

$$\Theta_2(v_1, v_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_2(x, y)\, e^{i[(x-a)v_1+(y-b)v_2]}\, dx\, dy. \tag{4.23}$$

Expanding the exponent in the integrand expression into an exponential series, and changing the order of summation and integration, we find

$$\Theta_2(v_1, v_2) = \sum_{k=0}^{\infty}\sum_{r=0}^{\infty} \frac{i^{k+r}}{k!\,r!} v_1^k v_2^r \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-a)^k (y-b)^r w_2(x, y)\, dx\, dy, \tag{4.24}$$

or

$$\Theta_2(v_1, v_2) = \sum_{k=0}^{\infty}\sum_{r=0}^{\infty} \frac{i^{k+r}}{k!\,r!} M_{kr}^{(k+r)} v_1^k v_2^r, \tag{4.25}$$

where $M_{kr}^{(k+r)}$ denotes mixed moments of the $(k + r)$-th order of the random variables ξ_j and η_j ($j = 1, 2, \ldots n$), which do not depend on j, since any pair of these random variables has the same distribution function $\omega_2(x, y)$.

The characteristic function for a normalized vector is equal to

$$\Theta_2\left(\frac{v_1}{\sigma_1\sqrt{n}}, \frac{v_2}{\sigma_2\sqrt{n}}\right) = \sum_{k=0}^{\infty}\sum_{r=0}^{\infty} \frac{i^{k+r}}{k!\,r!} \frac{M_{kr}^{(k+r)}}{\sigma_1^k \sigma_2^r n^{\frac{k+r}{2}}} v_1^k v_2^r. \tag{4.26}$$

The desired characteristic function of the sum of the normalized vectors (4.22) is, according to (3.77'), equal to the n-th power of the characteristic function (4.26), i.e.,

$$\begin{aligned} \Theta_{2n}(v_1, v_2) &= \left[\Theta_2\left(\frac{v_1}{\sigma_1\sqrt{n}}, \frac{v_2}{\sigma_2\sqrt{n}}\right)\right]^n = \\ &= \left[\sum_{k=0}^{\infty}\sum_{r=0}^{\infty} \frac{i^{k+r}}{k!\,r!} \frac{M_{kr}^{(k+r)}}{\sigma_1^k \sigma_2^r n^{\frac{k+r}{2}}} v_1^k v_2^r\right]^n = \left[1 - \frac{v_1^2 + 2rv_1v_2 + v_2^2}{2n} - \right. \\ &\quad \left. - \frac{i}{n^{3/2}}\left(\frac{\mathrm{k}_1 v_1^3 + \mathrm{k}_2 v_2^3}{6} + \frac{k_{12}v_1v_2^2 + k_{21}v_1^2v_2}{2}\right) + \ldots\right]^n. \end{aligned} \tag{4.27}$$

and of a sum of uniformly distributed terms, converge towards the normal law more rapidly than a sum of random variables with a Rayleigh law of distribution, although at first glance it seems as if the Rayleigh distribution were closer to the normal than the uniform distribution or that of (3.16). This is because the Rayleigh distribution is asymmetric, while the other two are symmetrical distributions.

4. Two-Dimensional Central Limiting Theorem.

The central limiting theorem of probability theory can also be expanded to multidimensional cases.

As has been indicated above, an aggregate of random variables $\xi_1, \xi_2, \ldots \xi_n$ may be treated as components of a vector in an n-dimensional space.

We shall examine the sum of the mutually independent vectors having identical n-dimensional distribution functions $w_n(x_1, x_2, \ldots x_n)$ of their components. Then, with certain limiting distributions, the component of the resultant vector (i. e., the sum of the vectors) approaches the n-dimensional normal law of distribution.

We shall examine in detail the two-dimensional case of this theorem, which figures in many applications, with the purpose of providing at the same time a correction to the normal law for the distribution function of a sum of mutually independent vectors in a plane, in relation to the number of n summed-up vectors and their numerical characteristics.

Let $(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_n, \eta_n)$ be the components of independent vectors in a plane, subject to the same two-dimensional law of distribution $\omega_2(x,y)$. We find the distribution function of the resultant of these vectors, the components (ξ, η) of which are equal to

$$\xi = \sum_{r=1}^{n} \xi_r, \quad \eta = \sum_{r=1}^{n} \eta_r. \tag{4.21}$$

For the solution of the problem at hand there will, as in the one-dimensional case, be employed the method of characteristic functions, there being first also determined the distribution function of the sum of normalized vectors with coordinates of

$$\bar{\xi} = \sum_{r=1}^{n} \frac{\xi_r - a}{\sigma_1 \sqrt{n}}, \quad \bar{\eta} = \sum_{r=1}^{n} \frac{\eta_r - b}{\sigma_2 \sqrt{n}}, \tag{4.22}$$

where $\underline{r}$ is the coefficient of correlation between the random variables ξ_j and η_j, $\underline{k}_1$ and $\underline{k}_2$ are the coefficients of asymmetry of each of the variables ξ_j and η_j respectively. Symbols $\underline{k}_{12}$ and $\underline{k}_{21}$ in (4.27) denote the following integrals:

$$k_{12}=\frac{1}{\sigma_1\sigma_2^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a)(y-b)^2 w_2(x, y)\,dx\,dy, \tag{4.28}$$

$$k_{21}=\frac{1}{\sigma_2\sigma_1^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a)^2(y-b) w_2(x, y)\,dx\,dy. \tag{4.29}$$

To investigate the behavior of the function $\Theta_{2n}(v_1, v_2)$ with large values of $\underline{n}$, we examine

$$\ln\Theta_{2n}(v_1, v_2)=n\ln\left[1-\frac{v_1^2+2rv_1v_2+v_2^2}{2n}-\right.$$
$$\left.-\frac{i}{n^{3/2}}\left(\frac{k_1v_1^3+k_2v_2^3}{6}+\frac{k_{12}v_1v_2^2+k_{21}v_1^2v_2}{2}\right)+\dots\right].$$

Expanding the logarithm into a series, we obtain

$$\ln\Theta_{2n}(v_1, v_2)=-\frac{v_1^2+2rv_1v_2+v_2^2}{2}-$$
$$-\frac{i}{\sqrt{n}}\left(\frac{k_1v_1^3+k_2v_2^3}{6}+\frac{k_{12}v_1v_2^2+k_{21}v_1^2v_2}{2}\right)+\dots \tag{4.30}$$

As the number, $\underline{n}$, of the terms increases without limit, we have

$$\lim_{n\to\infty}\Theta_{2n}(v_1, v_2)=e^{-\frac{1}{2}(v_1^2+2rv_1v_2+v_2^2)}, \tag{4.31}$$

which coincides with the characteristic function of the normalized deviations of two random variables, subject to the two-dimensional normal law of distribution (comp. 3.95).

In this manner it has been proved that the probability distribution of the components of a sum of normalized, mutually independent vectors, subject to the same law of distribution, tends towards the two-dimensional, normal distribution, as the number of terms is increased, regardless of what had been the distribution of the

components of the added vectors.

Having an expression of the two-dimensional distribution function of the components of a normalized resultant vector

$$w_{2n}(x, y) \sim \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}, \tag{4.32}$$

it is not difficult to find the asymptotic expression for the distribution function of a non-normalized resultant vector

$$W_{2n}(x, y) \sim \frac{1}{2\pi n\sigma_1\sigma_2\sqrt{1-r^2}} \times$$
$$\times e^{-\frac{1}{2n(1-r^2)}\left\{\frac{(x-na)^2}{\sigma_1^2} - \frac{2r(x-na)(y-nb)}{\sigma_1\sigma_2} + \frac{(y-nb)^2}{\sigma_2^2}\right\}}. \tag{4.32'}$$

Employing (4.30) it is possible, with a margin of error on the order of $1/n$, to establish a correction for the two-dimensional normal law of distribution (4.32). Since from (4.30) it follows, within the indicated margin of error, that

$$\Theta_{2n}(v_1, v_2) = e^{-\frac{1}{2}(v_1^2+2rv_1v_2+v_2^2)}\left[1 - \frac{i}{\sqrt{n}}\left(\frac{k_1v_1^3+k_2v_2^3}{6} + \right.\right.$$
$$\left.\left. + \frac{k_{12}v_1v_2^2+k_{21}v_1^2v_2}{2}\right)\right],$$

therefore, subjecting $\Theta_{2n}(v_1, v_2)$ to an inverse, two-dimensional Fourier transformation, we find the following expression for the distribution function $\omega_{2n}(x,y)$ of the components of the resultant vector $(\bar{\xi}, \bar{\eta})$

$$w_{2n}(x, y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(v_1x+v_2y)} e^{-\frac{1}{2}(v_1^2+2rv_1v_2+v_2^2)} \times$$
$$\times\left[1 - \frac{i}{\sqrt{n}}\left(\frac{k_1v_1^3+k_2v_2^3}{6} + \frac{k_{12}v_1v_2^2+k_{21}v_1^2v_2}{2}\right)\right] dv_1dv_2.$$

Breaking down the integral of the sum into a sum of integrals, and employing relationships analogous to (4.16)-(4.18), we find

$$w_{2n}(x, y) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x^2-2rxy+y^2}{2(1-r^2)}}\left[1 - \frac{1}{6\sqrt{n}\,(1-r^2)^3} \times\right.$$
$$\times\{(k_1 - rk_{12})[(ry-x)^3 - 3(1-r^2)(ry-x)] +$$
$$(k_2 - rk_{21})[(rx-y)^3 - 3(1-r^2)(rx-y)] + 3k_{12}y(1-r^2)\times$$
$$\left.[(ry-x)^2 - (1-r^2)] + 3k_{21}x(1-r^2)[(rx-y)^2 - (1-r^2)]\}\right]. \tag{4.33}$$

5. The "Random Walk" Problem

The two-dimensional, central limiting theorem of probability theory, proved in the preceding section, is closely linked to the famous problem of the random walk, encountered in many applications (reflection of radar signal from clouds, multibeam propagation in radio-relay communication (fading), Brownian movement, direction of plant growth, etc.).

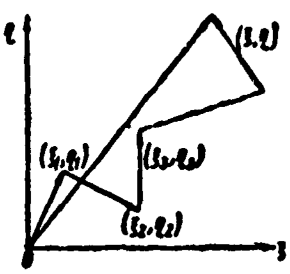


Fig. 30. Random migration of a point in a plane.

Let us examine a point which can move about in a plane, consecutively traversing linear segments, as shown in Fig. 30. The lengths of these segments and the directions of the moves are random, and are subject to a certain two-dimensional distribution function. The problem of the random walk consists of determining the probability density $W_{2n}(x,y)$ of the fact that, starting from the origin of the coordinates, the point after n consecutive, mutually independent moves will have the coordinates (x,y). It is obvious that any move may be treated as a vector with the random components (ξ_k, η_k); then the position of the point after n moves will be indicated by the end of the resultant vector (ξ, η), the components of which are equal to the summations (4.21).

For physical applications it is useful to regard the indicated resultant of the vectors as the sum of n isoperiodic vibrations with a given distribution function $\omega_2(r,\varphi)$, amplitudes $\rho_k = \sqrt{\xi_k^2 + \eta_k^2}$ and phases $\psi_k = \operatorname{arc\,tg} \frac{\eta_k}{\xi_k}$ ($k = 1, 2, \ldots n$). In such a form the problem of random moves is a generalization of the problem examined in Section 11 of Ch. 3, where it was assumed that the amplitudes of the vibrations

were constant, and that the phases were uniformly distributed between 0 and 2π.

One of the sources of interference in the reception of a radar signal is reflections from such objects as clouds, metallic strips, vegetation, etc. The mathematical description of the reflections of a radar signal from these objects is based on the assumption that they are caused by reflections from a large number of independent reflectors. A problem important in its applications lies in determining the distribution function of the amplitudes of the resultant signal, obtained from the addition of all these reflected vibrations. If it is assumed that the amplitude distribution of the reflected signals depends only on the distance r to the reflector and is equal to $\omega(r)$ for any of the signals, and the phases of these high-frequency vibrations are equiprobable and do not depend on the amplitudes, then the problem at hand may be solved exactly*.

Identifying each of the reflected signals with a vector, the length of which is $\rho_K=\sqrt{\xi_K^2+\eta_K^2}$, and whose direction is equiprobable, we return to the problem of the preceding paragraph.

Employing (3.29), we find the two-dimensional distribution function of the amplitude and phase of any of the reflected signals (vectors with components of ξ_k and η_k)

$$W_2(r,\varphi)=\frac{1}{2\pi}w(r)=rw_2(x,y). \tag{4.34}$$

The two-dimensional characteristic function of any of the vectors (ξ_k, η_k) is equal to

$$\Theta_2(v_1,v_2)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}w_2(x,y)\,e^{i(xv_1+yv_2)}\,dx\,dy=$$
$$=\frac{1}{2\pi}\int\limits_0^{\infty}\int\limits_0^{2\pi}w(r)\,e^{ir(v_1\cos\varphi+v_2\sin\varphi)}\,dr\,d\varphi.$$

Integration on φ, as in Section 5 of Ch. 3, leads to a zero-order Bessel function, i.e.,

$$\Theta_2(v_1,v_2)=\int\limits_0^{\infty}w(r)J_0\left(r\sqrt{v_1^2+v_2^2}\right)dr. \tag{4.35}$$

* This assumption is valid, if the distance from the radar receiver to the group of reflection sources is much greater than the distances between the sources. Moreover, the present problem does not consider the directivity pattern of the antenna.

The characteristic function of the resultant vector is, in accordance with (3.77'), equal to the n-th power of characteristic function (4.35);

$$\Theta_{2n}(v_1, v_2) = \left(\int_0^\infty w(r) J_0\left(r\sqrt{v_1^2+v_2^2}\right)dr\right)^n =$$

$$= \underbrace{\int_0^\infty\int_0^\infty \ldots \int_0^\infty}_{n \text{ times}} w(r_1)w(r_2)\ldots w(r_n) J_0\left(r_1\sqrt{v_1^2+v_2^2}\right) \times \tag{4.36}$$

$$\times J_0\left(r_2\sqrt{v_1^2+v_2^2}\right)\ldots J_0\left(r_n\sqrt{v_1^2+v_2^2}\right) dr_1\, dr_2 \ldots dr_n.$$

From (4.36), by means of a two-dimensional inverse Fourier transformation we find the desired distribution function of the components of the resultant vector

$$w_{2n}(x, y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Theta_{2n}(v_1, v_2)\, e^{-i(v_1x+v_2y)}\, dv_1\, dv_2 =$$

$$= \frac{1}{4\pi^2}\underbrace{\int_0^\infty \ldots \int_0^\infty}_{n \text{ times}} w(r_1)\ldots w(r_n) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J_0\left(r_1\sqrt{v_1^2+v_2^2}\right)\ldots$$

$$\ldots J_0\left(r_n\sqrt{v_1^2+v_2^2}\right) e^{-i(v_1x+v_2y)}\, dv_1 dv_2 dr_1 \ldots dr_n.$$

By a substitution of the integration variables $v_1 = s\cos\psi$, $v_2 = s\sin\psi$, we reduce the expression $\omega_{2n}(x,y)$ to the form of

$$w_{2n}(x, y) = \frac{1}{4\pi^2}\underbrace{\int_0^\infty \ldots \int_0^\infty}_{n \text{ times}} w(r_1)\ldots w(r_n) \int_0^\infty\int_0^{2\pi} J_0(r_1 s)\ldots \tag{4.37}$$

$$\ldots J_0(r_n s)\, e^{-is(x\cos\psi + y\sin\psi)}\, s\,ds\,d\psi\, dr_1 \ldots dr_n.$$

If in the distribution function $\omega_{2n}(x,y)$ we now pass to the variables r and φ, then in accordance with (3.29) we find that

$$W_{2n}(r, \varphi) = r w_{2n}(r\cos\varphi, r\sin\varphi) = \frac{r}{4\pi^2}\underbrace{\int_0^\infty \ldots \int_0^\infty}_{n \text{ times}} w(r_1)\ldots$$

$$\ldots w(r_n)\int_0^\infty\int_0^{2\pi} J_0(r_1 s)\ldots J_0(r_n s) \times$$

$$\times e^{-isr\cos(\psi-\varphi)}\, s\,d\psi\,ds\,dr_1 \ldots dr_n.$$

Integration with respect to ψ yields a zero-order Bessel function, then

$$W_{2n}(r, \varphi)=\frac{r}{2\pi}\underbrace{\int\limits_0^\infty \ldots \int\limits_0^\infty}_{n+1 \text{ times}} w(r_1)\ldots w(r_n) J_0(r_1 s)\ldots$$
$$\ldots J_0(r_n s) J_0(rs)\, s ds\, dr_1 \ldots dr_n, \tag{4.38}$$

from which we obtain the distribution function of the amplitude of the resultant signal

$$W_{1n}(r)=\int\limits_0^{2\pi} W_{2n}(r,\varphi)\, d\varphi = r\underbrace{\int\limits_0^\infty \ldots \int\limits_0^\infty}_{n+1 \text{ times}} w(r_1)\ldots$$
$$\ldots w(r_n) J_0(r_1 s)\ldots J_0(r_n s) J_0(r, s)\, s ds\, dr_1 \ldots dr_n. \tag{4.39}$$

Considering that the variables $r_1, r_2, \ldots r_n$ in integral (4.39) are separable, we find

$$W_{1n}(r)=r\int\limits_0^\infty \left(\int\limits_0^\infty w(v) J_0(vs)\, dv\right)^n s J_0(rs)\, ds, \quad r>0. \tag{4.40}$$

The corresponding integral distribution function $F_{1n}(R)$, i.e., the probability that the amplitude of the resultant vector will not exceed R, is equal to

$$F_{1n}(R)=\int\limits_0^R\int\limits_0^\infty \left(\int\limits_0^\infty w(v) J_0(vs)\, dv\right)^n rs J_0(rs)\, ds dr.$$

Since

$$\int\limits_0^R rs J_0(rs)\, dr = R J_1(Rs),$$

then

$$F_{1n}(R)=R\int\limits_0^\infty J_1(Rs)\left(\int\limits_0^\infty w(v) J_0(vs)\, dv\right)^n ds. \tag{4.41}$$

If the length of the summed-up vectors is constant and equal to a, then $\omega(r) = \delta(r - a)$ and from (4.41) we find

$$F_{1n}(R)=R\int\limits_0^\infty J_1(Rs) J_0^n(as)\, ds. \tag{4.42}$$

The probability that the length of the resultant of the vectors (the length of each is equal to a, and its direction is random) will not exceed the length of one

term is, in this case, equal to

$$F_{1n}(a)=a\int_0^\infty J_1(as)J_0^n(as)\,ds=\int_\infty^0 J_0^n(as)\frac{dJ_0(as)}{ds}=\frac{1}{n+1},$$

a resultant obtained by Kliver (i.e., Cleaver?) [7].

It is of interest to examine a case where the number of n reflected signals is great, which actually occurs in practice (for instance, the number of reflected vibrations from drops of rain). On the basis of the two-dimensional limiting theorem it is possible to say in advance that the distribution of the vector components of the resultant signal (after an appropriate normalization of the terms) will be approximately normal. Let us find, employing (4.40), a correction to this normal law in relation to the number n and to the distribution moments $\omega(r)$ of the amplitudes of the reflected vibrations.

The distribution function of the amplitude of the normalized resultant vibration is obtained from (4.40) by the substitution of $\rho=\sqrt{\frac{2}{m_2 n}}\ r$*

$$w_{1n}(\rho)=\sqrt{\frac{m_2 n}{2}}\,\rho\int_0^\infty\left(\int_0^\infty w(v)J_0(vs)\,dv\right)^n sJ_0\left(s\rho\sqrt{\frac{m_2 n}{2}}\right)\times$$
$$\times ds\sqrt{\frac{m_2 n}{2}}=\rho\int_0^\infty\left(\int_0^\infty w(v)J_0\left(vz\sqrt{\frac{2}{m_2 n}}\right)dv\right)^n zJ_0(\rho z)\,dz. \qquad (4.43)$$

Expanding the Bessel function into an exponential series

$$J_0\left(vz\sqrt{\frac{2}{m_2 n}}\right)=\sum_{k=0}^\infty\frac{(-1)^k v^{2k}z^{2k}}{(k!)^2 2^k m_2^k n^k}$$

and changing the order of summation and integration, we obtain

$$\int_0^\infty w(v)J_0\left(vz\sqrt{\frac{2}{m_2 n}}\right)dv=\sum_{k=0}^\infty\frac{(-1)^k m_{2k}z^{2k}}{(k!)^2 2^k m_2^k n^k}=$$
$$=1-\frac{z^2}{2n}+\frac{m_4 z^4}{16m_2^2 n^2}-\dots,$$

---

* It is necessary here to consider that the Jacobian of the transformation from r to p is equal to $\frac{dr}{d\rho}=\sqrt{\frac{m_2 n}{2}}$.

where m_{2k} are the even-order initial moments of the distribution function $\omega(r)$ of the amplitudes of the reflected signals.

Let us write $L(z) = \left(1 - \frac{z^2}{2n} + \frac{m_4 z^4}{16 m_2^2 n^2} - \ldots\right)^n$, then

$$\ln L(z) = n \ln\left(1 - \frac{z^2}{2n} + \frac{m_4 z^4}{16 m_2^2 n^2} - \ldots\right) \sim$$
$$\sim n\left(-\frac{z^2}{2n} + \frac{m_4 z^4}{16 m_2^2 n^2} - \frac{z^4}{8n^2} + \ldots\right) =$$
$$= -\frac{z^2}{2} + \frac{z^4}{16n}\left(\frac{m_4}{m_2^2} - 2\right) + \ldots,$$

wherefrom

$$L(z) \sim e^{-\frac{z^2}{2}}\left[1 + \frac{z^4}{16n}\left(\frac{m_4}{m_2^2} - 2\right) + \ldots\right].$$

Substituting the obtained expansion of the integrand function into (4.43), we obtain with a precision of terms of the order of $\frac{1}{n^2}$ (cf. Appendix VI)

$$w_{1n}(\rho) = \rho \int_0^\infty z e^{-\frac{z^2}{2}} J_0(\rho z)\left[1 + \frac{z^4}{16n}\left(\frac{m_4}{m_2^2} - 2\right)\right] dz =$$
$$= \rho e^{-\frac{\rho^2}{2}} + \rho e^{-\frac{\rho^2}{2}} \cdot \frac{1}{2n}\left(\frac{m_4}{m_2^2} - 2\right) {}_1F_1\left(-2, 1, \frac{\rho^2}{2}\right),$$

and since

$${}_1F_1\left(-2, 1, \frac{\rho^2}{2}\right) = \left(1 - \rho^2 + \frac{\rho^4}{8}\right),$$

then

$$w_{1n}(\rho) = \rho e^{-\frac{\rho^2}{2}}\left\{1 + \frac{1}{16n}\left(\frac{m_4}{m_2^2} - 2\right)[\rho^4 - 8(\rho^2 - 1)]\right\}. \tag{4.44}$$

From (4.44) we find the desired distribution function of the amplitude of the resultant signal, when the number of n reflections is large:

$$W_{1n}(r) = \frac{2r}{m_2 n} e^{-\frac{r^2}{m_2 n}}\left[1 + \frac{1}{2n}\left(\frac{m_4}{m_2^2} - 2\right)\left(\frac{r^4}{2m_2^2 n^2} - \frac{2r^2}{m_2 n} + 1\right)\right]. \tag{4.45}$$

Thus, with a large number of reflections, the amplitude of the resultant signal is distributed according to the Rayleigh law. This result corresponds fully to the two-dimensional central limiting theorem, since it indicates that the components of

the resultant are normally distributed.

Let us note that the distribution function $\frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$ was first obtained by Rayleigh precisely as a law of the distribution of the resultant amplitude of the sum of a large number of isoperiodic vibrations with random phases.

If the components ξ_k and η_k of the reflected signals are independent and are normally distributed with zero means and dispersions of $\frac{1}{2}m_2$, the distribution $\omega(r)$ of the amplitudes of these signals will be a Rayleigh one (cf. Section 5, Ch. 2)

$$w(r) = \frac{2r}{m_2} e^{-\frac{r^2}{m_2}}.$$

Then the distribution of the amplitude of the sum of n independent reflected signals will, in accordance with (4.40), be equal to

$$W_{1n}(r) = r \int_0^\infty \left(\int_0^\infty \frac{2v}{m_2} e^{-\frac{v^2}{m_2}} J_0(vs)\, dv \right)^n s J_0(rs)\, ds =$$
$$= r \int_0^\infty e^{-\frac{nm_2}{4} s^2} s J_0(rs)\, ds = \frac{2r}{nm_2} e^{-\frac{r^2}{nm_2}},$$

i.e., precisely a Rayleigh distribution with a dispersion of nm_2, equal to the sum of the dispersions of the amplitudes of the terms. Since for the Rayleigh law of distribution $m_4 = 2 m_2^2$, the correction itself, depending on n in formula (4.45), turns in this case to zero.

The solution presented here of the problem of the amplitude distribution of a resultant may also be obtained by the general method, which is used for proving the two-dimensional central limiting theorem, i.e., by the resolution into a series of the characteristic function (4.35)

6. Law of Large Numbers

Let there be measured a certain unknown physical quantity a. For increased precision and reliability several measurements are usually made, and then the arithmetic mean of the measurements is supposed to equal the quantity a. The validity of such a supposition is based on the law of large numbers.

The results of measurements, which are made independently of one another, may be regarded as the random variables $\xi_1, \xi_2, \ldots \xi_n$. The mean value of each of

these random variables is equal to the quantity a being measured, i.e.,

$$m_1\{\xi_k\} = a. \tag{4.46}$$

This signifies that the measurements are free of systematic errors. If the measurements are made with a guarantee of any (albeit small) precision, the dispersions of the random variables ξ_k are limited

$$M_2\{\xi_k\} \leq \sigma^2. \tag{4.47}$$

Let us examine the arithmetic mean of the measurements, i.e., the arithmetic mean of the mutually independent random variables $\xi_1, \xi_2, \ldots \xi_n$

$$\xi = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}. \tag{4.48}$$

The mean value of the arithmetic mean of ξ is equal to

$$m_1\{\xi\} = \frac{1}{n}\sum_{k=1}^{n} m_1\{\xi_k\} = a, \tag{4.49}$$

and the dispersion is

$$M_2\{\xi\} = \frac{1}{n^2}\sum_{k=1}^{n} M_2\{\xi_k\} \leq \frac{\sigma^2}{n}. \tag{4.50}$$

It follows from (4.50), that the dispersion of the arithmetic mean of a sum is not less than n times smaller than the greatest of the dispersions of the terms, which indicates a reduction in the scattering of the values of the arithmetic mean with an increase of n.

The law of large numbers asserts that with a sufficiently large number of measurements, it is possible, with a probability close to unity, to consider that the arithmetic mean of the measurement results will be as close as is desired to the true value of the quantity, a, being measured.

The validity of the law of large numbers may be expected from the inequality

(4.50), its rigorous proof being based on one inequality, credited to Chebyshev.

Let a be the mean value of the random variable ξ, the distribution function of which is $\omega(x)$. Let us evaluate the probability of the fact that the random variable ξ deviates from its mean value, a, at least by the arbitrary positive number ε (Fig. 31, shaded region)

$$P\{|\xi - a| \geqslant \varepsilon\} = \int_{|x-a| \geqslant \varepsilon} w(x)\,dx,$$

and since $\frac{(x-a)^2}{\xi^2} \geqslant 1$, then

$$P\{|\xi - a| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} \int_{|x-a| \geqslant \varepsilon} (x-a)^2 w(x)\,dx.$$

If the limits of integration are expanded to extend from $-\infty$ to $+\infty$, then, bearing in mind that $\omega(x) \geqslant 0$, the size of the integral cannot diminish; then

$$P\{|\xi - a| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x-a)^2 w(x)\,dx = \frac{M_2\{\xi\}}{\varepsilon^2}. \qquad (4.51)$$

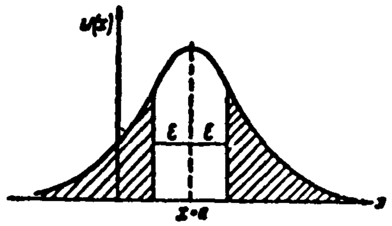


Fig. 31. Probability of the deviation of a random variable from its mean value by not less than ε is equal to the shaded area.

From (4.51) it follows that $P\{|\xi - a| \geqslant K\sqrt{M_2}\} \leqslant \frac{1}{K^2}$, i.e., deviations of ξ from its mean, considerably exceeding the mean-square, are improbable.

Employing the Chebyshev inequality (4.51), it is possible to evaluate the probability that the arithmetic mean ξ of the sum of n mutually independent random variables deviates from the mean value of each of the terms by more than the arbitrary amount $\varepsilon > 0$. Since in accordance with (4.50) the dispersion of the arith-

metic mean does not exceed $\frac{\sigma^2}{n}$, then

$$P\{|\xi - a| \geqslant \varepsilon\} < \frac{\sigma^2}{n\varepsilon^2}. \qquad (4.52)$$

If the quantities σ^2 and ε are fixed, then from (4.52) there follows the limiting relationship

$$\lim_{n\to\infty} P\{|\xi - a| \geqslant \varepsilon\} = 0. \qquad (4.53)$$

The probability limit of the converse inequality is equal to unity

$$\lim_{n\to\infty} P\{|\xi - a| \leqslant \varepsilon\} = 1. \qquad (4.54)$$

Formulas (4.53) and (4.54) are an analytic statement of the law of large numbers, which may be formulated in the following manner: if the random variables ξ_1, $\xi_2, \ldots \xi_n$ are mutually independent, have the same mean values and terminal dispersions, then the probability of the declination of the arithmetic mean of these random variables from their mean value by an amount greater than $\varepsilon>0$, tends toward zero when $n \longrightarrow \infty$, no matter how small ε be.

While an individual random variable may take values very far from its mean value, it is practically certain that the arithmetic mean of a large number of independent random variables takes values close to the mean value of the random variable. This is explained by the fact that in the formation of an arithmetic mean, the random deviation of the individual terms of a sum from their mean value cancel each other out to a considerable extent.

7. Simplest Problem of the Theory of Errors

The law of large numbers, like the Lyapunov theorem, is one to those relationships which are valid for a very large number of summed-up independent random variables. For the practical application of these limiting relationships it is necessary to have an estimate of the error in relation to the number of n items. A similar estimate was made for the central limiting theorem in Section 2.

According to the law of large numbers with a very large n, it is practically

certain that

$$\xi = \frac{\xi_1 + \xi_2 + \dots \xi_n}{n} \approx a, \tag{4.55}$$

where ξ_k (k = 1, 2, ...n) are mutually independent random variables which have the same mean values equal to a.

Let us make it our purpose, with a given small number of n measurements to find the probability of the fact that the deviation of the true value of the quantity a, undergoing measurement, from the arithmetic mean ξ of the experimental data will not exceed a certain fixed $\varepsilon > 0$, i.e.,

$$P\{-\varepsilon < \xi - a < \varepsilon\} = P\{\xi - \varepsilon < a < \xi + \varepsilon\} = \alpha. \tag{4.56}$$

The probability α is called the reliability, and the quantity ε, the precision of the approximate equality (4.55) with a given number of n measurements. The interval ($\xi - \varepsilon$, $\xi + \varepsilon$) which contains the true value of the quantity a being measured, is called the confidence interval.

The simplest problem of the theory of errors consists in estimating the precision (the length of the confidence interval) and the reliability of the approximate equality (4.55) for a small number of n measurements.

If the precision ε is given (i.e., the length of the confidence interval), then according to the law of large numbers the reliability of the measurements will grow as the number of measurements increases, approaching unity. With a given number of measurements the reliability will be the greater, the wider is the confidence interval, i.e., the greater is the permissible error.

In other words, with a given number of measurements it is impossible to increase reliability without reducing precision or, conversely, to increase precision without reducing reliability.

The objective of the error theory is to find a link between the quantities of reliability α, precision ε, and the number of measurements n, which would make it possible on the basis of a given two of these quantities to find the third.

The overall error of physical measurement may be regarded as the sum of a large number of mutually independent elementary errors. Therefore the overall error must according to the Lyapunov theorem be normally distributed. The assumption that the measurement results, ξ_k, of the quantity, a, are subject to the normal law of distribution

$$w(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-a)^2}{2\sigma^2}}, \tag{4.57}$$

is the basis of error theory.

If the dispersion σ^2 of the measurement results were known a priori, the question of estimating the reliability and precision would be solved very simply. For instance, it could be asserted with practical certainty, i.e., with a probability of $\alpha = 0.997$, that the true value of the quantity being measured differs from the arithmetic mean by no more than 3σ.

Let us take, as the approximate value of the dispersion of any of the random variables, ξ_k, the quantity

$$\sigma^2 \approx \frac{1}{n-1}\sum_{k=1}^{n}(\xi_k-\bar{\xi})^2. \tag{4.58}$$

Then the approximate value of the dispersion of the arithmetic mean of the independent random variables ξ_k will be

$$\eta^2=\frac{\sigma^2}{n}=\frac{1}{n(n-1)}\sum_{k=1}^{n}(\xi_k-\bar{\xi})^2. \tag{4.59}$$

The random variable η^2 will, in accordance with the results of Section 9 of Ch.3, an χ^2 type distribution.

It is natural to suppose that the deviation $\bar{\xi}-a$ is proportional to the root of the dispersion η^2, i.e., $\bar{\xi}-a = \eta t$.

The random variable

$$t=\frac{\bar{\xi}-a}{\eta} \tag{4.60}$$

is the normalized error of measurement. This random variable is equal to the ratio of two random variables, one of which is distributed normally (with a zero mean), while the distribution of the other is linked with the χ^2 distribution.

Employing the formulas of Section 3.Ch.3, it is possible to show that when $n \gg 2$ the distribution function $S_n(x)$ of the random variable is equal to

$$S_n(x) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}. \tag{4.61}$$

The distribution given by function (4.61) is known by the name of the Student distribution or the t-distribution.

For large values of n, the distribution of normalized error, in full accordance with the Lyapunov theorem, approaches the normal

$$\lim_{n\to\infty} S_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{4.62}$$

However, for small values of n, the Student distribution differs markedly from the limiting normal one. Fig. 32 shows the curve of distribution (4.61) for $n = 4$ and the curve of the limiting normal distribution (4.62). It follows from the equations of these curves that the probabilities of a large deviation are greater for the random variable than for a normally distributed random variable.

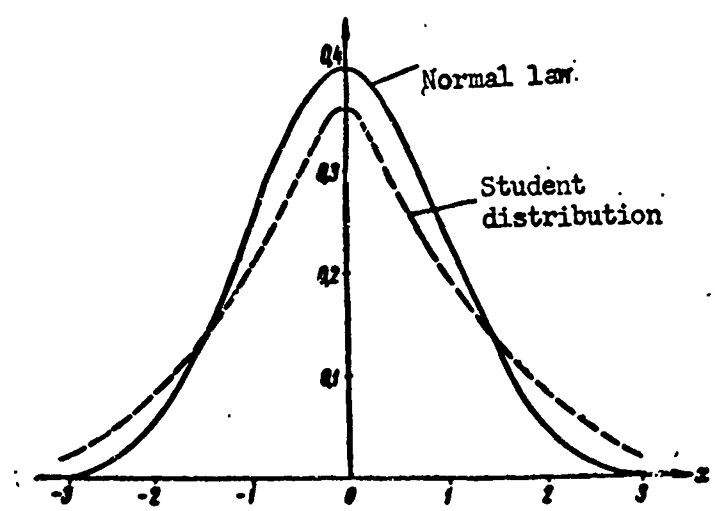


Fig. 32. Probability density of normal distribution and of student distribution when $n = 4$

The integral distribution function, i.e., the probability that t will not exceed the quantity x, is equal to

$$P(t < x) = \int_{-\infty}^{x} S_n(x)\,dx. \tag{4.63}$$

When $n \longrightarrow \infty$, as follows from (4.62), the integral $\int_{-\infty}^{x} S_n(x)dx \longrightarrow F(x)$. Therefore the Laplace function $F(x)$ is sometimes called the <u>integral of error probability.</u>

Returning to the relationship (4.56), we introduce the quantity t_α, proportional to an error of ε with a given reliability of α

$$\varepsilon = \eta t_\alpha. \tag{4.64}$$

Then the inequality

$$-\varepsilon < \xi - a < \varepsilon \tag{4.65}$$

corresponds uniquely to the inequality

$$-t_\alpha < t < t_\alpha. \tag{4.66}$$

Therefore the probability, that the quantity a being measured lies within the confidence interval (4.65), is equal in accordance with (4.56) to

$$P\{\xi - \varepsilon < a < \xi + \varepsilon\} = P\{\xi - t_\alpha \eta < a < \xi + t_\alpha \eta\} =$$
$$= P\{-t_\alpha < t < t_\alpha\} = \int_{-t_\alpha}^{t_\alpha} S_n(x)\,dx = \alpha,$$

and, since the Student function is even,

$$P\{\xi - \varepsilon < a < \xi + \varepsilon\} = 2\int_{0}^{t_\alpha} S_n(x)dx = \alpha. \tag{4.67}$$

Equality (4.67) establishes the desired dependence between the quantities α, ε, and n, i.e., between reliability, precision, and the number of measurements.

Appendix III contains a table of the values of t_α which satisfy (4.67) for given instances of α and n. The last line of the table ($n = \infty$) corresponds to a normal distribution of error.

The magnitude, t_α, is the proportionality factor between the length of the confidence interval (precision), and the mean-square deviation of the results of measurement.

With a given reliability α the smallest t_α (i.e., the smallest length of the confidence interval) corresponds to a normal distribution of error. Thus, for instance in Section 3 of Ch. 2 it was pointed out that for a normally distributed random variable it is practically certain (with a probability of 0.997) that none of its values will deviate by more tha 3η from the mean, i.e., that the length of the confidence interval is equal to 6η. Actually, particularly with a small number of measurements, the length of this interval, corresponding to a practical certainty, will always be greater.

8. Determination of Distance by Independent Observers

As an illustration of the statistical method, presented above, of processing measurement results let us examine an example connected with determining the distance of a certain object by means of radar. To increase reliability, the indicator is observed by several independent observers, and then the arithmetic mean of the observed distances is taken as the true value of the distance.

Let there be the following five distance measurements, made by five independent observers:

$\xi_1 = 39.1$ km; $\xi_2 = 39.7$ km; $\xi_3 = 38.3$ km; $\xi_4 = 39.6$ km; $\xi_5 = 38.1$ km.

The arithmetic mean of these observations yields the distance of

$$\xi = \frac{39{,}1 + 39{,}7 + 38{,}3 + 39{,}6 + 38{,}8}{5} = 39{,}1 \text{ km.}$$

The approximate dispersion of the arithmetic mean of the observations is, in accordance with (4.59), equal to

$$\eta^2 = \frac{1}{4 \cdot 5}(0^2 + 0{,}6^2 + 0{,}8^2 + 0{,}5^2 + 0{,}3^2) = \frac{1{,}34}{20} = 0{,}067,$$

or

$$\eta = 0.26 \text{ km.}$$

Let us assign a reliability of $\alpha = 0.99$ and estimate the precision of the distance determination. In the table in Appendix III, for $\alpha = 0.99$ and $n = 5$ we find that $t_{\alpha} = 4.604$, i.e., the confidence interval extends in each direction from the arithmetic mean for an extent equal to almost five mean-square errors. Consequently, it is possible to assert with a probability of 0.99 that the true distance from the object lies within the limits of from $\xi - t_{\alpha}\eta = 39.1 + 4.604 \cdot 0.26 = 37.91$ km to $\xi + t_{\alpha}\eta = 39.1 + 4.604 \cdot 0.26 = 40.29$ km, i.e., the distance is determined with an error not greater than $\varepsilon = 1.19$ km, which constitutes 3%. In order to have a narrower confidence interval with the same reliability, five observers are insufficient.

If normal distribution had been used for measuring the error, instead of the Student distribution, it would have been practically certain that the true distance lies within the limits of $\xi - 3\eta = 39.1 - 3 \cdot 0.26 = 38.32$ km to $\xi + 3\eta = 39.1 + 3 \cdot 0.26 = 39.88$ km, which corresponds to an error of only 780 m or 2%.

It is possible to solve a problem of another type: with an assigned permissible error of, say 0.5%, to find the probability that the distance, found as the arithmetic mean of 5 independent observations, will differ from the true distance by no more than by 195 m (which constitutes about 0.5% of the distance measured.) Then $t_{\alpha} = \frac{\xi}{\eta} = \frac{0.005 \cdot 39.1}{0.26} = 0.75$, and by the table in Appendix III, after linear extrapolation, we find for n = 5 the desired probability of $\alpha = 0.504$.

Thus, from the results of five independent measurements it is impossible, with an error not exceeding 195 m (0.5%), to form a sufficiently reliable judgment as to the distance measured, while for precision on the order of 1 km (3%), this number of observations was entirely sufficient.

Literature

1. A. Ya. Khinchin. Matematicheskiye osnovaniya kvantovoy statistiki (The Mathematical Foundations of Quantum Statistics). Gostekhizdat, 1951.
2. S. Chandrasekar. Stokhasticheskiye problemy v fizike i astronomii (Stochastic Problems in Physics and Astronomy). For. Lit. Pub. Hse., 1947.

3. V. I. Romanovskiy. Primeneniye matematicheskoy statistiki v opytnom dele (The Employment of Mathematical Statistics in Experimentation). Gostekhizdat, 1947.

4. V. I. Romanovskiy. Osnovnyye zadachi teorii oshibok (Basic Problems of Error Theory). Gostekhizdat, 1947.

5. B. V. Gnedenko, A. N. Kolmogorov. Predel'nyye raspredeleniya dlya summ nezavisimykh sluchaynykh velichin (Limiting Distribution for Sums of Independent Random Variables), 1949 /Place not specified/.

6. K. A. Norton, L. E. Vogeler, W. J. Mansfield, P. J. Short. The Probability Distribution of the Amplitude of a Constant Vector Plus a Rayleigh-Distributed Vector. Proc. IRE, v. 43, No. 10, 1955.

7. B. Van der Pol and H. Bremmer. Operatisionnoye ischisleniye na osnove dvukhstoronnego preobrazovaniya Laplasa (Operational Calculus Based on the Two-sided Laplace Integral) For. Lit. Pub. Hse., 1952. (New York, 1950.)

8. V. I. Siforov. Nakopleniye shumoy izamiraniy v magistralnykh radioreleynykh liniyakh (Noise and Fading Accumulation in Arterial Radio Relay Systems.) "Elektrosvyaz'" ("Electric Communications", No. 5, May 1956.

Chapter V

RANDOM PROCESSES

1. Definitions and Terminology

The preceding chapters presented the basic rules of classical probability theory, as well as examples illustrating the applications of this theory to the solution of certain technical problems. A large part of this theory is devoted to the study of the laws of distribution of one random variable or of a system of a finite number of random variables. The following chapters will be devoted to the exposition of certain results, and applications to radio engineering, of the modern theory of random processes — of a theory which deals with random variables which depend on a continuously changing parameter, time.

The best-known example of a random process, to which in the course of our presentation we shall be constantly returning, are fluctuation (shot and thermal) noises in radio devices (both of the transmitting and the receiving kinds). Thus, in the observation of noise voltage at the output of identical instruments, it appears that the forms of this voltage, or the time functions describing this random process, differ. The task of the theory of random processes lies in finding the statistical laws governing the various functions which describe the same physical phenomenon (fluctuation noise), analogously to the manner in which the classical probability theory establishes statistical laws by the observation of random variables.

Quantitatively, a random process is described as a random function of time $\xi(t)$, the values of which are random variables at any moment in time.

Just as the random variable ξ is characterized by an aggregate of possible values and by a distribution of the probabilities that these values will occur so the random process $\xi(t)$ (a random function of time) is defined as an aggregate of time functions, together with the laws characterizing the statistical properties of this aggregate. Each of the functions of this aggregate is called the realization of a random function.

The realizations of the random function $\xi(t)$ are designated by $\xi^{(1)}(t)$, $\xi^{(2)}(t)$,.... The results of various observations of the same process will be quantitatively defined as various realizations of the corresponding random function.

In order to characterize the statistical properties of a random function, it will be necessary to generalize the concept of probability.

Let us examine N realizations of a random function. From this number let us isolate those n_1 realizations, whose values in a specified time instant, $t = t_1$, are smaller than a certain number x_1. It is possible to express the following assumption, based on experiment: with a sufficiently large number of N, the relative portion $\frac{n_1}{N}$ of the functions which in the moment of time $t = t_1$ lie below x_1, will possess statistical stability, i.e., will remain an approximately constant number. This number is called the probability that, when $t = t_1$, the random function $\xi(t)$ lies below x_1; it is designated by $P\{\xi(t_1) < x_1\}$.

The indicated probability will depend both on the fixed moment of time and on the selected level, x_1, i.e., it will be a function of the two variables t_1 and x_1. This function

$$F_1(x_1, t_1) = P\{\xi(t_1) \leqslant x_1\} \tag{5.1}$$

is called the <u>first integral probability distribution function</u> of a random process. If the integral distribution function has a partial derivative with respect to x_1,

$$\frac{\partial F_1(x_1, t_1)}{\partial x_1} = w_1(x_1, t_1), \tag{5.2}$$

this derivative is called the <u>probability density</u> or the first <u>(one-dimensional) distribution function</u> of a random process.

Functions $F_1(x_1, t_1)$ and $w_1(x_1, t_1)$ are the simplest statistical characteristics of a random process. These characteristics provide a concept of the process only in separate, fixed time instants. It can be said that they characterize the process statistically and do not give a concept of the dynamics of its development.

For a more complete characterization of a random process it is necessary to know the connection between the probable values of a random function for the two arbitrary

time instants, t_1 and t_2. For this we again examine N realizations of a random function and isolate from this number those n_2 realizations, the values of which at the time instant $t = t_1$ are smaller than x_1, and at the time instant $t = t_2$ are smaller than x_2.

It is possible to assume, analogously to the preceding case, that with a sufficiently large N, the relative portion $\frac{n_2}{N}$ of the functions at $t = t_1$ lying below the level of x_1, and at $t = t_2$ below the level of x_2, will possess statistical stability, i.e., will remain an approximately constant number. This number is called the probability that at $t = t_1$ the random function $\xi(t)$ lies below the level of x_1, and at $t = t_2$ below the level of x_2.

The indicated probability $P\{\xi(t_1) \leqslant x_1, \xi(t_2) \leqslant x_2\}$ is a function of the four variables x_1, x_2, t_1, t_2

$$F_2(x_1, x_2, t_1, t_2) = P\{\xi(t_1) \leqslant x_1, \xi(t_2) \leqslant x_2\} \quad (5.3)$$

and is called the second integral probability distribution function of a random process.

If the function $F_2(x_1, x_2, t_1, t_2)$ has a derivative

$$\frac{\partial^2 F_2(x_1, x_2, t_1, t_2)}{\partial x_1 \partial x_2} = w_2(x_1, x_2, t_1, t_2), \quad (5.4)$$

this derivative is called the two-dimensional probability density or the second distribution function. Two-dimensional distribution functions are statistical characteristics of the random process, on the basis of which it is possible to evaluate the relationship between the probable values of a random function at two arbitrary time instants.

In a similar manner it is also possible to determine the probability that the random function $\xi(t)$ will, at n instants of $t_1, t_2, \ldots t_n$, lie below the levels of $x_1, x_2, \ldots x_n$ respectively

$$P\{\xi(t_1) \leqslant x_1;\ \xi(t_2) \leqslant x_2;\ \ldots\ \xi(t_n) \leqslant x_n\} = \\ = F_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n). \quad (5.5)$$

The indicated probability is a function of 2 n variables and is called the n-th integral probability distribution function of a random process. If $F_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n)$ has a derivative

$$\frac{\partial^n F_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n)}{\partial x_1 \partial x_2 \ldots \partial x_n} = w_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n), \tag{5.6}$$

this derivative is called the n-dimensional probability density or the n-th distribution function.

The sequence of distribution functions $w_1(x_1, t_1)$, $w_2(x_1, x_2, t_1, t_2)$, ... $w_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n)$ constitute a unique ladder, in the ascent of which more detailed familiarity with the random process is acquired*. Knowing the first n distribution functions, it is possible to evaluate the relationship between the probable values of a random function at n arbitrary time instants. The distribution-function sequence in question must possess all the properties of probability distribution functions.

Specifically, from a distribution function of the n-th order it is possible to obtain all the distribution functions of the lower order, through to the first [cf. (2.47)].

Thus, the statistical properties of a random process (of a random function) may be characterized by means of an n-dimensional distribution function, and the more accurately so the greater is the number of n.

Limiting ourselves to the n-dimensional distribution function of a random process, we essentially identify the random function with the aggregate of n random variables $\xi(t_1), \xi(t_2), \ldots \xi(t_n)$ — values of the random function $\xi(t)$ at the fixed time instants $t_1, t_2, \ldots t_n$.

---

* Of course, the statistical properties of a random process may be described by the sequence of characteristic functions $\Theta_1(\nu_1, t_1)$, $\Theta_2(\nu_1, \nu_2, t_1, t_2)$, ..., linked by Fourier transformations to the distribution functions $w_1(x_1, t_1)$, $w_2(x_1, x_2, t_1, t_2)$, ... (cf. Section 7, Ch.III).

For a graphic illustration of the concepts introduced [here] , we turn to a geometric representation of an aggregate of realizations of the random function $\xi(t)$ in the form of the set of plane curves $\xi^{(1)}(t)$, $\xi^{(2)}(t)$,... $\xi^{(k)}(t)$, ... (Fig. 33).

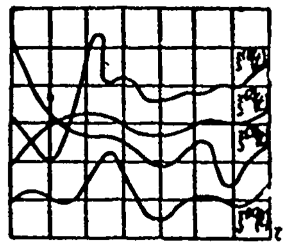


Fig. 33. Realization aggregate of a random function.

Let us now replace each of the $\xi^{(k)}(t)$ curves by a sequence of rectangular pulses of the same width T, equal to the period of their repetition. The amplitude of any m-th pulse of this sequence is selected as being equal (Fig. 34) to

$$A_{mk}=\xi^{(k)}\left(mT-\frac{T}{2}\right).$$

Let us call the sequence fixed in this manner, a sequence of base pulses*.

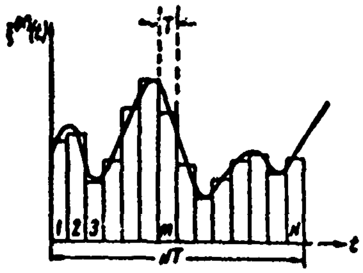


Fig. 34. Sequence of base pulses.

* The replacement of a process realization by a base pulse sequence is known in communications practice as the time quantization (breaking-up into parts) of a signal.

If the curve $\xi^{(k)}(t)$ is reproduced by a radio system with a pass band Δ, then, if the width of the base pulse satisfies the inequality

$$T \ll \frac{1}{\Delta},$$

the reproduction of $\xi^{(k)}(t)$ by the system is practically equivalent to the reproduction of the base pulse sequence, since in a system with the indicated pass band it is impossible to distinguish a change in the $\xi^{(k)}(t)$ curve within an interval of time shorter than T. Replacing each of the realizations $\xi^{(k)}(t)$ by a sequence of base pulses, we obtain an aggregate of sequences of rectangular pulses occurring in instants of time, representing multiples of their duration, the amplitude of an impulse occurring at the instant mT, being capable of taking one of the values of $A_{m1}, A_{m2}, \ldots A_{mk}, \ldots$.

An aggregate of these amplitude values may be treated as the random variable $\xi(t_m)$, which characterizes the random amplitude of a pulse occurring at the instant $t_m = mT$.

Thus the random process $\xi(t)$ is approximately represented in the form of a sequence of base pulses with a random amplitude of ξ_m, this approximation being the more precise, the shorter the duration of the base pulse.

It follows from what has been said, that a one-dimensional distribution function of a random process represents the probability density of the random amplitudes of base pulses, which may depend on the number of the pulse (i.e., on time). A two-dimensional distribution function of a random process yields the joint distribution of the random amplitudes of any pair of base pulses. Analogously, an n-dimensional distribution function yields the joint distribution of any group consisting of n base pulses.

The simplest type of a random process will be one in which the random amplitudes of the base pulses are mutually independent. Such a random process is fully characterized by a one-dimensional distribution function. In the general case, the amplitudes of the base pulses occurring at different points of time will be dependent, and for a complete characterization of a group of n base pulses it will be necessary to

assign an n-dimensional distribution function to their random amplitudes.

Very frequently, the random variable $\xi(t)$ may contain, as an item, a given function of time S(t)

$$\xi(t)=\xi_1(t)+S(t). \tag{5.7}$$

Let us agree to say that $\xi_1(t)$ and S(t) represent respectively, the purely random and the determined parts of process $\xi(t)$. It is clear that the indicated parts of a process are always independent.

Since the probability density for S(t) is represented by the delta-function $\delta[x - S(t)]$ (cf. Section 3, Ch. II), the distribution function of sum (5.7) is, in accordance with (3.26) (cf. also Appendix IV)

$$W_1(x, t)=\int_{-\infty}^{\infty} w_1(u, t)\delta[x-S(t)-u]\,du = w_1[x-S(t), t], \tag{5.8}$$

where $w_1(x,t)$ is the distribution function of the purely random part of the process.

Analogously, the two-dimensional distribution function of sum (5.7) is

$$W_2(x_1, x_2, t_1, t_2)=w_2[x_1-S(t_1), x_2-S(t_2), t_1, t_2], \tag{5.8'}$$

where $w_2(x_1, x_2, t_1, t_2)$ is the two-dimensional distribution function of $\xi_1(t)$.

In many radio-engineering problems, the subjects of investigation are the useful signals and the accompanying interference, the aggregates of which constitute a random process. The signal is often regarded as the determined part of this process, and the interference as the purely random part. However, in the modern general theory of communications (information theory), the signals are regarded not as given functions of time, but as an aggregate of possible functions of time, possessing definite statistical characteristics, i.e., as a purely random process. On the other hand, problems are possible where interference (for instance, artificially created) becomes the determined part of the process. The expressions for the distribution

functions of signals jointly with noise will depend on what assumptions are made concerning the nature of each of the indicated processes.

2. Stationarity

We now pass to a presentation of the properties of the most important class of random processes — the so-called stationary random processes.

The random process $\xi(t)$ is called stationary if its distribution function $w_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n)$, of an arbitrary n order, does not change with any shift of the entire group of points $t_1, t_2, \ldots t_n$ along the time axis, i.e., if for any n and τ

$$w_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n) = = w_n(x_1, x_2, \ldots x_n, t_1+\tau, t_2+\tau, \ldots t_n+\tau). \quad (5.9)$$

In other words, a random process will be stationary when the expressions for a distribution function of any order do not depend on the position of the starting point of the time reckoning.

It follows from this definition, that for a stationary random process:

a) the first distribution function has the same form at any instant in time (i.e., does not depend on time)

$$w_1(x, t) = w_1(x, t+\tau) = w_1(x), \quad (5.10)$$

b) the second distribution function may depend only on the difference $t_2 - t_1$.

$$w_2(x_1, x_2, t_1, t_2) = w_2(x_1, x_2, t_2 - t_1), \quad (5.11)$$

c) the third distribution function may depend only on the two differences $t_3 - t_1$ and $t_2 - t_1$

$$w_3(x_1, x_2, x_3, t_1, t_2, t_3) = w_3(x_1, x_2, x_3, t_3 - t_1, t_2 - t_1) \quad (5.12)$$

etc.

In the representation of a random process by a sequence of base pulses the above

properties of a stationary process will signify that: 1) the amplitude distribution functions of all base pulses are the same; 2) the joint distribution of the random amplitudes of any pair of base pulses depends only on the distance between the pulses, and does not depend on the location of this pair on the time axis; 3) the joint distribution of any triplet of base pulses depends only on the distances between one pulse and the two others, and does not depend on the location of these pulses on the time axis, etc.

In a number of problems, it suffices to know less about a random process than is disclosed by the distribution functions: it is, for instance, possible to limit one's self to such summary numerical characteristics of a process as the initial moments of distribution, which in the general case, are determined by means of (2.96) in the following manner:

$$m_{k_1 k_2 \ldots k_n}(t_1, t_2, \ldots t_n) = m_1\{[\xi(t_1)]^{k_1}[\xi(t_2)]^{k_2}\ldots[\xi(t_n)]^{k_n}\} = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n \text{ times}} x_1^{k_1}x_2^{k_2}\ldots x_n^{k_n} \times w_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n)\,dx_1 dx_2 \ldots dx_n. \tag{5.13}$$

These distribution moments may also be determined with the aid of characteristic functions by means of formula (3.78).

The simplest numerical characteristics of a random process are the distribution moments of the first two orders:

The mean value of the random process (or the first moment of the one-dimensional law of distribution)

$$m_1\{\xi(t)\} = \int_{-\infty}^{\infty} x w_1(x, t)\,dx = a(t); \tag{5.14}$$

the dispersion of the random process (or the second central moment of the one-dimensional law of distribution)

$$m_1\{[\xi(t) - a]^2\} = \int_{-\infty}^{\infty}(x - a)^2 w_1(x, t)\,dx = \sigma^2(t); \tag{5.15}$$

the correlation function of the random process (or the mixed, second initial moment of the two-dimension law of distribution)

$$m_1\{\xi(t_1)\xi(t_2)\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2w_2(x_1,\ x_2,\ t_1,\ t_2)\ dx_1dx_2=\\ =B(t_1,\ t_2). \qquad (5.16)$$

If the two-dimensional characteristic function $\Theta_2(v_1, v_2, t_1, t_2)$ of a random process is known, the correlation function can be found, in accordance with (3.78), by means of the formula

$$B(t_1,\ t_2)=-\left(\frac{\partial^2\Theta_2}{\partial v_1\partial v_2}\right)_{v_1=0,\ v_2=0}. \qquad (5.16')$$

It follows from (5.10) that, for a stationary random process, its mean value a and dispersion σ^2 are constants, not depending on time.

Since the second distribution function $w_2(x_1, x_2, t_2 - t_1)$ of a random stationary process depends only on the difference $\tau = t_2 - t_1$, therefore the correlation function of such a process also depends only on the variable τ

$$B(\tau)=B(t_2-t_1)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2w_2(x_1,\ x_2,\ t_2-t_1)\,dx_1dx_2. \qquad (5.17)$$

While the experimental determination of the two-dimensional distribution function of a random process constitutes a cumbersome and in practice rarely achievable task, the experimental determination of the correlation function of a stationary random process can, as will be shown below, be accomplished with comparative ease. For a very large number of problems the correlation function is a sufficiently complete statistical characteristic of a stationary random process.

The part of theory devoted to the study of only those properties of random processes, which are determined by their moments of the first two orders, is called the correlational theory of random processes. Since multidimensional distributions have no place in correlational theory, it is natural within the framework of this theory to consider as stationary all random processes whose mean value and dispersion

do not depend on time, and whose correlation function depends only on the time difference of $\tau = t_2 - t_1$. Random processes satisfying these conditions are called stationary in the broad sense (or stationary in the sense of A. Ya. Khinshin).

Stationarity in the broad sense is not identical with the precise definition of stationarity in accordance with (5.9). Random processes which are stationary according to (5.9) will, of course, always be stationary in the broad sense, but not vice versa.

Within the framework of the correlational theory, it is clearly sufficient to know the random-process distribution function of an order no higher than the second.

It should be noted, that the two-dimensional distribution function, and even more so, the correlation function, cannot describe a random process in as much detail as can multi-dimensional distribution functions. Moreover, the same correlation function can correspond to different processes, in other words, an equality of correlation functions does not signify an identity of processes. However, at present only the correlational theory of random processes has been developed to the extent that a widespread application of it has become possible in various engineering problems. The practical value of the correlational theory is increased by virtue of the fact that a large class of stationary random processes — that of the so-called normal random processes — is fully defined by the correlation function alone. For these processes the concepts of stationarity in the precise and in the broad senses coincide.

The random processes represented by random functions of a continuously changing argument t, which are frequently encountered in radio-engineering problems, do not, however, completely exhaust the demands of practical application.

Sometimes a random process is thought of as broken down into consecutive steps, i.e., is represented by random functions of a discrete time t. Such a process was the sequence of base pulses, whose amplitudes represented the aggregate of the random variables which depend on the time instant t_j (m = -2, -1, 0, 1, 2,...). Another example of a random process with discrete time is the simple Markov chain (cf. Section 10,

Ch. 1). This chain possesses the special property, that the probability of a transition from the state of $x(s)$ to one of the possible states of ξ when $t = s + 1$ does not depend on what states occurred at the instants, $t < s$. Random processes possessing the indicated property are called processes without aftereffect. The absence of an aftereffect indicates that the statistical interrelationships in a random process with discrete time do not extend further than for one step.

If in the Markov chains the limitation of the values of t to whole numbers is discarded, we obtain the general Markov process, directed by the probability of the transition of $P\left\{t_0, x, t_1\right\}$ from the state of x at the instants t_0, to one of the aggregate of states of ξ at the instant t. The general Markov process is also a random process with discrete time without aftereffect.

An exposition of the theory of Markov processes goes beyond the bounds of the present book. In the future we shall deal primarily with such random processes with continuous time, for which a statistical link between past states and future ones is essential.

3. Ergodicity of Stationary Processes

Certain random processes possess a property important to practical applications of the theory under discussion, namely ergodicity. Without going into the fine points of strict definitions, let us note that a random process is ergodic, if any of its realizations has the same statistical properties, i.e., if with the passage of time a given realization undergoes on the average the same changes that can affect any other realization of the random process. Therefore, in the case of ergodic random processes, any statistical characteristic obtained by the averaging of an aggregate of possible realizations can, with a probability of as close as desired to unity, be obtained by the averaging a single realization of a random process over a sufficiently large interval.

Let us observe over a sufficiently long period of time, T, some realization $\xi^{(k)}(t)$ of a random process (for instance, the noise voltage at the output of some

receiving set). Let, for the time T, the total time that the realization (the noise voltage) stays within the definite limits of from u_1 to u_2 (Fig. 35) be equal to

$$t^{(k)}(T)=\Delta t_1+\Delta t_2+\ldots+\Delta t_m.$$

The ratio $\nu_k(T)=\frac{t^{(k)}(T)}{T}$ is called the relative time of dwell of realization $\xi^{(k)}(t)$ within the limits of from u_1 to u_2.

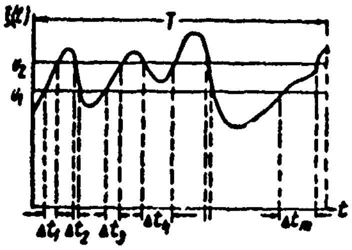


Fig. 35. Total time of dwell of a realizati n within the limits of u_1, u_2.

Let us, for example, examine the position of a material point, moving in accordance with the law of $\eta = a\sin\omega\xi$, at a random time instant $t=\xi$. The relative time of dwell of this point below the level of $x(-a\leqslant x<a)$ is equal (Fig. 36) to

$$F(x)=\frac{t_1+\left(\frac{2\pi}{\omega}-t_2\right)}{\frac{2\pi}{\omega}}=\frac{\frac{1}{\omega}\arcsin\frac{x}{a}+\frac{2\pi}{\omega}-\left(\frac{\pi}{\omega}-\frac{1}{\omega}\arcsin\frac{x}{a}\right)}{\frac{2\pi}{\omega}},$$

or

$$F(x)=\frac{1}{2}+\frac{1}{\pi}\arcsin\frac{x}{a},\quad |x|\leqslant a.$$

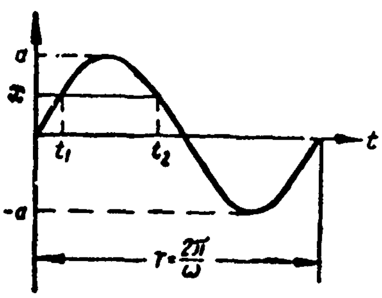


Fig. 36. Position, at a random time instant, of a point moving according to harmonic law.

In the example under examination the relative time of dwell simply coincides with the integral distribution function of the random variable η [compare (3.17)].

If the random process, $\xi(t)$ is ergodic, given large values of T it is possible to consider that the relative time of dwell $\nu_K(T)$, coincides with the probability $P\{u_1 < \xi(t) \leqslant u_2\}$ of the fact that the values of a random function are bounded by the limits of $u_1 < \xi(t) \leqslant u_2$. This probability has been defined above as the share of the realizations satisfying a given condition. Since $\nu_K(T)$ is a random variable, its proxim' y to the constant P, at large values of t may be asserted only with some (sufficiently great) probability. We have already encountered a similar circumstance in Chapter IV (the law of large numbers), where it was shown that the arithmetic mean of a large number of independent random measurements may, with a probability of close to unity, be considered as the mean value a of a random variable (which is obtained by averaging the aggregate of its possible values $a = \int_{-\infty}^{+\infty} xw(x)\,dx$). Therefore, the equivalence of the two methods of averaging ergodic random processes — averaging an aggregate of realizations and the averaging of a single realization with respect to time — is also sometimes called the law of large numbers.

The soviet mathematician A. Ya. Khinchin has proved an important theorem to the effect that stationary random processes (with sufficiently general assumptions [2]) are ergodic. From this ergodic theorem there directly follows the possibility, in the computation of such important characteristics of the stationary random process as the distribution moments of $m_k\{\xi(t)\}$ and the correlation function $B(\tau) = m_k \{\xi(t)\xi(t+\tau)\}$ (which are averages for an aggregate of realizations), of restricting one's self to only one single realization of the random process, namely

$$m_k\{\xi(t)\} = m_1\{[\xi(t)]^k\} = \lim_{T\to\infty} \frac{1}{T}\int_0^T [\xi^{(r)}(t)]^k dt, \tag{5.18}$$

$$B(\tau) = m_1\{\xi(t)\xi(t+\tau)\} = \lim_{T\to\infty} \frac{1}{T}\int_0^T \xi^{(r)}(t)\xi^r(t+\tau)\,dt \tag{5.19}$$

independently of r, i.e., of which of the realizations is being averaged with respect to time. An averaging of the product $\xi(t)\,\xi(t+\tau)$ with respect to time is called the autocorrelation function of process $\xi(t)$.

From (5.18), when k = 1, we find the mean value of a stationary random process

$$a=\lim_{T\to\infty}\frac{1}{T}\int_0^T \xi^{(r)}(t)\,dt, \tag{5.20}$$

which may thus be treated as a constant component of this process.

The second-order initial moment equal to the value of the correlation function, when $\tau = 0$, is obtained from (5.18), when k = 2

$$m_2=B(0)=\lim_{T\to\infty}\frac{1}{T}\int_0^T [\xi^{(r)}(t)]^2 dt. \tag{5.21}$$

If $\xi(t)$ represents a change in a voltage or of a current, regarded as a stationary random process, then, as can be seen from (5.21), the quantity m_2 = B(0) is equal to the average power developed by the process at a load of 1 ohm.

The limit in the right part of (5.21) yields the square of the effective value of the voltage and, consequently, establishes the equivalence of the square of the effective value to the quantity B(0) (or to the dispersion of the random process, if its mean value is equal to zero).

The ergodic theorem is not only of theoretical, but also of considerable practical interest. From the equivalence of the two methods of averaging stationary random processes, it follows that within the framework of the correlational theory there is no necessity of studying a large aggregate of realizations, which the researcher as a rule does not have at his disposal, but there suffices one single realization, observable over an extended period of time. Thus, for example, the statistical properties of fluctuation noises, which constitute a stationary random process, are studied in the course of a sufficiently long time on one radio receiver, and then the results of this investigation may be extended to all identical sets. Since actual observations are limited in time, the equality of the aggregate averages with time averages may be accepted with but a definite degree of certainty and precision. This circumstance is emphasized in Table 6 in which the formulas of the two methods for averaging random variables and stationary random processes are compared.

Table 6

Random variable	Stationary random process
Average of aggregate of possible values $$a = m_1\{\xi\} = \int_{-\infty}^{\infty} xw(x)\,dx$$	Average of aggregate of realizations $$a = m_1\{\xi(t)\} = \int_{-\infty}^{\infty} xw_1(x)\,dx,$$ $$B(\tau) = m_1\{\xi(t)\xi(t+\tau)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2w_2(x_1,x_2,\tau)\,dx_1dx_2$$
Arithmetic mean of independent observations $$\xi = \frac{\xi_1+\xi_2+\ldots+\xi_n}{n}$$	Time average of one realization $$\varphi_r(T) = \frac{1}{T}\int_0^T \xi^{(r)}(t)\,dt$$ $$\psi_r(T,\tau) = \frac{1}{T}\int_0^T \xi^{(r)}(t)\xi^{(r)}(t+\tau)\,dt$$
Law of large numbers $$\lim_{n\to\infty} P\{\mid\xi-a\mid<\varepsilon\} = 1$$	Ergodic theorem $$\lim_{T\to\infty} P\{\mid\varphi_r(T)-a\mid<\varepsilon\} = 1,$$ $$\lim_{T\to\infty} P\{\mid\psi_r(T,\tau)-B(\tau)\mid<\varepsilon\} = 1$$

For practical calculations of average values and of correlation functions of stationary random processes from experimental curves of $\xi^{(r)}(t)$ it is necessary in formulas (5.19) and (5.20) to replace the integrals by the integral sums

$$a \approx \frac{1}{N}\sum_{k=1}^{N} \xi^{(r)}(k\delta)$$

and

$$B(\tau) \approx \frac{1}{N}\sum_{k=1}^{N} \xi^{(r)}(k\delta)\xi^{(r)}(k\delta+\tau).$$

where δ is sufficiently small, so that over the period of time δ it would not be necessary to allow for changes in the curve of $\xi^{(r)}(t)$, and N is selected to make $T = N\delta$ so great that an additional increase in the number of items in the sum would have little effect on the result of the summation. In recent years special calculating machines — correlators — have appeared, which considerably facilitate the laborious calculation of correlation function from experimental curves.

The basic elements of every correlator are the delay circuit, the multiplier, the integrating circuit, and the recording apparatus. Depending whether the multiplication is performed by the digitol or analogue methods, the correlators are classified (similarly to computers) as digitol or analogue types. The former are more precise, but are considerably more complex in design. The latter are simpler and in many practical cases provide satisfactory precision. Integration (accumulation) may be accomplished, as is well known, by means of a simple RC circuit. A more rigorous examination of the integration process in the RC integrator, especially necessary in the investigation of short-term correlation, makes it necessary to introduce under the sign of the correlation integral still another function, which takes into account the characteristic of the RC circuit, i.e., makes it necessary to replace the ordinary correlation function by the so-called short-term auto-correlation function [16]

$$B(\tau, t) = 2\alpha \int_{-\infty}^{t} \xi(x)\,\xi(x+\tau)\, e^{-2\alpha(t-x)}\,dx,$$

where $\alpha = \frac{1}{2RC}$

Let us note that in the ergodic theorem the stationarity of a process is understood in the precise sense. In order to emphasize the importance of this remark, let us examine, for instance, the determined periodic process S(t) with a period of T. Since the aggregate of the realizations of the process consists of the one function S(t), the average for the aggregate coincides with this function, i.e., depends on time, while the average over time is a direct component of the process. Thus the average for the aggregate does not coincide in this case with the time average. It

is not difficult, however, to show that the autocorrelation function of a periodic process depends only on the temporal displacement $\tau = t_2 - t_1$. In actual fact, representing the periodic function by the Fourier series

$$S(t)=\sum_{n=0}^{\infty} c_n \cos\left(\frac{2\pi n}{T} t+\varphi_n\right),$$

we obtain (by virtue of the periodicity, the average is taken for a period)

$$B(\tau)=\frac{1}{T}\int_0^T \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} c_n c_k \cos\left(\frac{2\pi n}{T} t+\varphi_n\right)\cos\left[\frac{2\pi k}{T}(t+\tau)+\varphi_k\right]dt.$$

But in view of the orthogonality of the trigonometric functions, the integrals of the product of the cosines when $n \neq k$, are equal to zero, and when $n = k \neq 0$

$$\frac{1}{T}\int_0^T \cos\left(\frac{2\pi n}{T} t+\varphi_n\right)\cos\left[\frac{2\pi n}{T}(t+\tau)+\varphi_n\right]dt=\frac{1}{2}\cos\frac{2\pi n}{T}\tau.$$

Thus, for a periodic process, the autocorrelation function is

$$B(\tau)=\sum_{n=0}^{\infty} \varepsilon_n \frac{c_n^2}{2}\cos\frac{2\pi n}{T}\tau, \qquad (5.22)$$

where $\varepsilon_n = 2$, when $n = 0$, and $\varepsilon_n = 1$, when $n \geqslant 1$.

It follows from (5.22) that the autocorrelation function of a periodic process is a periodic function with the period of the process in question, the amplitudes of the harmonics of the fundamental frequency being obtained from the corresponding amplitude of the harmonic of the periodic process by means of squaring and dividing by two. The autocorrelation function does not depend at all on the phase angles which contain harmonics of the initial periodic process.

4. Properties of the Correlation Function

Let us examine first of all the behavior of a correlation function with an unlimited increase of the argument τ. If a random function does not contain determined (periodic) components (a purely random process), the dependence between the variables $\xi(t+\tau)$ and $\xi(t)$ must weaken when $\tau \to \infty$ and these variables become independent at the limit. Since the mean value of a product of independent random variables is equal to the product of the mean values of the comultiples, and since for a station-

ary process the mean value does not depend on time, then

$$\lim_{\tau\to\infty} B(\tau) = B(\infty) = \lim_{\tau\to\infty} m_1\{\xi(t)\xi(t+\tau)\} = a^2 \tag{5.23}$$

or

$$a = \sqrt{B(\infty)}. \tag{5.24}$$

Thus the mean value of a stationary random process is equal to the square root of the asymptotic value of the correlation function when $\tau \to \infty$.

On the other hand,

$$\lim_{\tau\to 0} B(\tau) = B(0) = m_1\{[\xi(t)]^2\} = m_2\{\xi(t)\},$$

and since $\sigma^2 = m_2 - a^2$, and the dispersion of a stationary process does not depend on time, then

$$\sigma^2 = B(0) - a^2,$$

or, taking into account (5.24)

$$\sigma^2 = B(0) - B(\infty). \tag{5.25}$$

Thus, the dispersion of a stationary random process is equal to the difference of the values of the correlation function when $\tau = 0$ and $\tau = \infty$, i.e., to the difference between the average power of the process and the power of the direct component.

In view of the stationarity, i.e., of the independence of the distribution functions from the origin of the time reckoning, the correlation function $B(\tau)$ is an even function.

$$B(\tau) = B(-\tau). \tag{5.26}$$

Let us further examine $m_1\{[\xi(t) \pm \xi(t+\tau)]^2\}$ which, as the mean value of an essentially positive quantity, cannot be negative. Since

$$m_1\{[\xi(t) \pm \xi(t+\tau)]^2\} = m_1\{\xi^2(t) \pm 2\xi(t)\xi(t+\tau) + \xi^2(t+\tau)\} =$$
$$= 2B(0) \pm 2B(\tau) \geqslant 0,$$

therefore

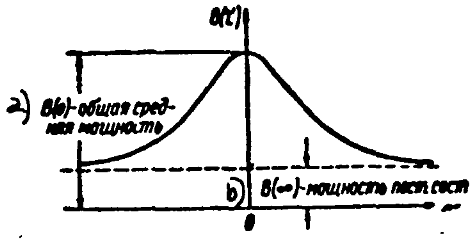
$$B(0) \geq |B(\tau)|. \tag{5.27}$$

Thus, no value of the correlation function can exceed the value of this function when $\tau = 0$.

This affirmation follows also from (5.19) and (5.21), since

$$\int_0^T \xi^{(r)}(t)\xi^{(r)}(t+\tau)\,dt \leq \int_0^T [\xi^{(r)}(t)]^2 dt.$$

Figure 37 shows a typical curve of the correlation function of a random process, which illustrates the above enumerated properties of this function. It should however, be noted that the asymptotic approach of $B(\tau)$ to the quantity a^2 when $\tau \to \infty$ does not always take place monotonously; there may be cases when the values of the correlation function fluctuate about a^2, approaching this quantity with an increase of τ.



Legend:

a) B(o)-general mean power
b) B(∞)-power of the constant component

Fig. 37. Typical curve of correlation function of a purely random stationary process.

Let us remark that equality (5.19) established for stationary random processes the actual physical sense of the correlation function as a function which characterizes the connection between the preceding and succeeding values of ξ (t). As a measure of this connection we shall take the amount $2\varepsilon(\tau)$ of the mean-square deviation of $\xi^{(k)}(t)$ from this same realization, but shifted by an interval τ along the time axis (Fig. 38), i.e.,

$$2\varepsilon(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T [\xi^{(k)}(t) - \xi^{(k)}(t+\tau)]^2 dt.$$

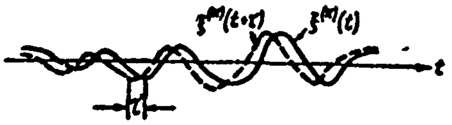


Fig. 38. Time shift of the realization of a random process.

Squaring the difference of the functions, breaking the integral of the sums into a sum of integrals and employing the ergodic theorem, we obtain

$$\varepsilon(\tau) = B(0) - B(\tau),$$

or

$$B(\tau) + \varepsilon(\tau) = B(0) = \text{const}.$$

Consequently, the correlation function $B(\tau)$ supplements the magnitude of the mean square deviation, $\xi^{(K)}(t+\tau)$, from $\xi^{(K)}(t)$, to a constant equal to the power of the stationary random process (Fig. 39), and indicates in this manner the extent to which, on the average two values of a random process, separated by a time interval of τ are linked to each other.

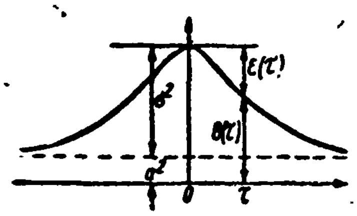


Fig. 39. Link between $\varepsilon(\tau)$ and $B(\tau)$.

In place of a random process, there are frequently examined the deviations from its mean value called the pulsations of a random process, or fluctuations*

$$\xi_0(t) = \xi(t) - a.$$

---

* Sometimes a fluctuation is called the numerical characteristic of a random variable - the square root of the dispersion (cf. p. 65). It should always be kept in mind, in which of these two completely different senses this term is employed.

The correlation function $B_0(\tau)$ of the pulsations of a stationary random process is equal to

$$B_0(\tau)=m_1\{\xi_0(t)\xi_0(t+\tau)\}=m_1\{\xi(t)\xi(t+\tau)-a\xi(t)-\\-a\xi(t+\tau)+a^2\}, \qquad (5.28)$$

and since the mean value of the sum is equal to the sum of the mean values, therefore

$$B_0(\tau)=B(\tau)-a^2. \qquad (5.29)$$

It follows from (5.29) that the mean value of the pulsation is equal to zero, and its dispersion to

$$\sigma^2=B_0(0). \qquad (5.30)$$

The ratio

$$R(\tau)=\frac{B_0(\tau)}{B_0(0)}=\frac{B_0(\tau)}{\sigma^2} \qquad (5.31)$$

is called the coefficient of correlation of a stationary random process (or of a stationary random process with a zero mean value). The quantity $R(\tau)$ is sometimes called the normalized correlation function. This term is more precise, but the term "coefficient of correlation" has become entrenched in the literature.

Comparing (5.31) with (2.91), we conclude that $R(\tau)$ is the coefficient of correlation of the two random variables $\xi(t_1)$ and $\xi(t_2)$, with $t_2 = t_1 + \tau$.

The coefficient of correlation $R(\tau)$ has the same properties as the correlation function. It constitutes an even function of this argument. The maximum value corresponds to $\tau = 0$ and, according to (5.27) and (5.31), $R(0) = 1$, and $|R(\tau)| \leq 1$ for any τ. For a purely random process when $\tau \to \infty$, $R(\tau) \to 0$. The coefficient of correlation may assume values of zero even for a finite value of τ. However, the equality of the coefficient of the correlation does not yet signify independence, whereas two independent random variables are always uncoordinated, i.e., for them always $R = 0$ (cf. p. 74).

For a purely random process, it is always possible to indicate such a τ_0, that when $\tau > \tau_0$, the variables $\xi(t)$ and $\xi(t+\tau)$ may be considered practically independent, practical independence being understood in the sense that when $\tau > \tau_0$, the

absolute amount of the coefficient of correlation remains smaller than a given quantity, for instance

$$|R(\tau)| < 0{,}05. \qquad (5.31) \text{ [sic]}$$

The quantity τ_0 is called the time or the interval of correlation.

Sometimes the time of correlation τ_0 is defined as half the width of the base of a rectangle of unit height, whose area is equal to the area under the correlation coefficient curve, i.e.,

$$\tau_0 = \frac{1}{2}\int_{-\infty}^{\infty} R(\tau)\,d\tau = \frac{1}{B(0)}\int_0^{\infty} B(\tau)\,d\tau. \qquad (5.32)$$

In the future, when not the correlation function of a process, but the coefficient of correlation is discussed, it is necessary to remember that the latter refers to the pulsation of a process or to a process with a zero mean value.

5. Detection of a Periodic Signal in Noise

The so-called correlation method of receiving a periodic signal concealed in noise is based on the employment of the properties of the correlation function.

Let $\xi(t)$ be the stationary random process representing noise, and S(t) the periodic signal. Let us examine the autocorrelation function $B(\tau)$ of the sum of $\xi(t)$ and S(t). In accordance with (5.19) we have

$$B(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T [\xi(t) + S(t)][\xi(t+\tau) + S(t+\tau)]\,dt = \\ = B_{ш}(\tau) + B_{c}(\tau) + B_{шc}(\tau) + B_{cш}(\tau), \qquad (5.33)$$

with the designations of

$$B_{ш}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T \xi(t)\xi(t+\tau)\,dt, \qquad (5.34)$$

$$B_{c}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T S(t)S(t+\tau)\,dt, \qquad (5.35)$$

$$B_{шc}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_0^T \xi(t)S(t+\tau)\,dt, \qquad (5.36)$$

$$B_{cш}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_0^T S(t)\xi(t+\tau)\,dt. \qquad (5.37)$$

The functions $B_ш(\tau)$ and $B_c(\tau)$ represent respectively the autocorrelation functions of noise and signal. The functions $B_{шc}$ and $B_{cш}$ are called mutual correlation function of the processes $\xi(t)$ and $S(t)$. If the processes $\xi(t)$ and $S(t)$ are independent and the mean value of one of these processes is equal to zero, then $B_{шc} = B_{cш} \equiv 0$.

Processes for which the mutual correlation functions are equal to a constant or turn to zero are often called incoherent functions. In the general case, mutual correlation functions can serve as a measure of the link (coherence) between the two processes.

Since the signal and the noise at the input of a receiving set may, as a rule, be considered independent, the mutual correlation functions in (5.33) disappear and then

$$B(\tau)=B_ш(\tau)+B_c(\tau), \qquad (5.38)$$

i.e., the autocorrelation function of the sum of the noise and the periodic signal is equal to the sum of the autocorrelation functions of the noise and of the signal. As has been shown above, the correlation function of a random stationary process. such as the noise is, coincides with its autocorrelation function and diminishes with an unlimited increase of the argument. For a periodic signal the autocorrelation function is periodic, with a period equal to the period of the signal. Therefore, by switching a correlator into the receiving set, it is possible after some time to detect whether the correlation function of the received signal has a periodic component. If such a periodic component stands out when τ is increased, this will indicate the presence of a signal; if however with the growth of τ the value of the correlation function tends toward a constant value (in particular towards zero) then, consequently, a signal is lacking*.

* For short periods of the observation time T, all four terms of equation (5.33) will influence the findings of the correlator, but, with a sufficiently large time of integration, the terms corresponding to the mutual correlation of signal and noise will disappear.

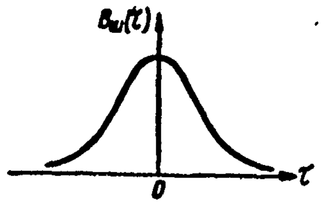
Figure 40 shows the autocorrelation functions of noise, of a periodic signal, and of their sum. The autocorrelation function of the sum has an absolute maximum when $\tau = 0$ and relative maximums which, starting with some τ, are repeated with a period practically equal to the period of the signal. If P_c is the power of the signal, and $P_{ш}$ is the power of the noises, then from (5.21) and (5.22) it follows that

$$B(0) = P_c + P_{ш}, \quad B(nT) = P_c$$

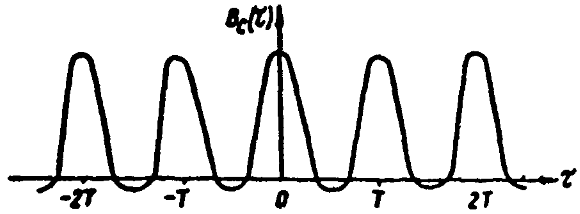
and, consequently, the relative difference between the absolute maximum and the periodically repeating maximums of the autocorrelation function

$$\frac{B(0) - B(nT)}{B(nT)} = \frac{P_{ш}}{P_c}$$

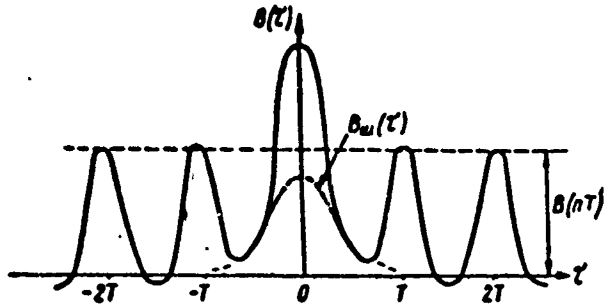
is smaller, the more the power of the signal exceeds the power of the noise. Therefore, with a strong signal, the detection of a periodic signal in noise may be accomplished sufficiently rapidly, whereas with a weak signal, in order to obtain the correlation function of a large argument τ, time will be required for the retention of the process for time τ, and for the averaging over the time of $nT \gg \tau$. Thus although theoretically no limitations exist for the detection of weak signals by the correlation method, in practice there will exist threshold values for signals detectable in noise, determined by a finite observation time and by the solving capacity of the instruments.



a)



b)



c)

Fig. 40. Correlation functions: a) of noise; b) of periodic signal; c) of sum of periodic signal and noise.

In cases when it is necessary to detect a periodic signal S(t), the form of which is known in advance, the correlation method of separating the signal from noise may be improved on the basis of an a priori knowledge of the function S(t)(coherent reception). For this, the signal S(t) is reproduced at the receiving end by a local oscillator, and there is studied the mutual correlation function of the incoming signal, distorted by noise, and of the signal from the local oscillator, i.e.,

$$B(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_0^T[\xi(t)+S(t)]\,S(t+\tau)\,dt=B_c(\tau)+B_{шc}(\tau).$$

Since $\xi(t)$ and S(t) are independent, therefore, as has already been pointed out above $B_{шc}(\tau) = 0$ and, consequently,

$$B(\tau)=B_c(\tau). \qquad (5.38')$$

Thus, in the case under consideration the autocorrelation function of the signal is directly obtained at the output of the correlator, and no supplementary time is needed for the removal of the aperiodic item in (5.38) which is caused by noise. Clearly in this case too, the remark made on page 182, concerning the influence of the term $B_{шc}(\tau)$ since the averaging time T is finite, remains in force.

Besides those indicated above, such methods of receiving weak signals as accumulation and filtration are at present known and in widespread use. The essence of all these methods is described mathematically by the averaging operation

$$B_T(\tau)=\frac{1}{T}\int_0^T[\xi(t)+S(t)]f(t+\tau)\,dt.$$

The function f(t) under the integral sign determines the method of reception. Thus, with the autocorrelation method, $f(t) = \xi(t) + S(t)$, and with the coherent method, $f(t) = S(t)$. Accumulation corresponds with $f(t) \equiv 1$, and filtration with $f(t) = h(t)$, where h(t) is the pulse transfer function of the filter (cf. Section 1, Ch.VI).

6. Power Spectrum of a Random Process

The extent to which harmonic analysis has been fruitfully employed in the study of non-random (determined) processes is well known: Fourier series for periodic, and the Fourier integral for aperiodic processes. It would be desirable to have such a simple and effective mathematical apparatus for the study of random processes. The direct application of classical harmonic analysis to random processes is impossible, since analytically these processes are not expressed as precise functional relationships of an independent variable. It is, however, possible to generalize harmonic analysis, averaging the spectral breakdowns obtained from individual realizations, with respect to time.

With this purpose, let us examine one realization $\xi^{(k)}(t)$ of a random process. Let in addition $\xi_T^{(k)}(t)$ be a function equal to zero outside the interval of $|t| \leqslant T/2$ and coinciding with $\xi^{(k)}(t)$ within this interval. The spectrum (Fourier transformation) of function $\xi_T^{(k)}(t)$ has the form of

$$Z_{kT}(i\omega)=\int_{-\frac{T}{2}}^{\frac{T}{2}}\xi_T^{(k)}(t)\,e^{-i\omega t}dt. \qquad (5.39)$$

However, for the function $\xi_T^{(k)}(t)$ bounded in time by the interval of $|t| \leqslant T/2$, the continuous spectrum $Z_{kt}(i\omega)$ is uniquely determined by the assignment of spectral intensities at the discrete points $\omega = \frac{2\pi k}{T}$, located at a distance of $\frac{2\pi}{T}$ from each other. Actually, resolving the function $\xi_T^{(k)}(t)$ into a Fourier series over the

interval of $-T/2 \leqslant t \leqslant T/2$, we obtain

$$\xi_T^{(k)}(t)=\sum_{n=-\infty}^{\infty} a_n^{(k)} e^{\frac{2\pi i n}{T} t}. \qquad (5.40)$$

where

$$a_n^{(k)}=\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \xi_T^{(k)}(t) e^{-\frac{2\pi i n}{T} t} dt=\frac{1}{T} Z_{kT}\left(\frac{2\pi i n}{T}\right). \qquad (5.40')$$

Substituting (5.40) into (5.39) and employing (5.40'), after the simplest computations we then obtain

$$Z_{kT}(i\omega)=\sum_{n=-\infty}^{\infty} Z_{kT}\left(\frac{2\pi i n}{T}\right)\frac{\sin\left(\frac{\omega T}{2}-\pi n\right)}{\frac{\omega T}{2}-\pi n}. \qquad (5.41)$$

Relationship (5.41) yields the indicated link between the continuous spectrum $Z_{kT}(i\omega)$ and its values at the discrete points of $\frac{2\pi n}{T}$*.

If $\xi_T^{(k)}$ is a voltage or current at a load of 1 ohm, then the mean power of a process at a frequency of $\frac{2\pi n}{T}$, with respect to the frequency band of $\Delta f = 1/T$, will be equal in accordance with (5.40) and (5.40') to

$$G_{kT}\left(\frac{2\pi n}{T}\right)=\frac{2}{\Delta f}|a_n^{(k)}|^2=\frac{2}{T}\left|Z_{kT}\left(\frac{2\pi i n}{T}\right)\right|^2.$$

Passing to the limit when $T \to \infty$, we obtain

$$G_k(\omega)=\lim_{T\to\infty}\frac{2}{T}|Z_{kT}(i\omega)|^2,$$

which has the dimensions of power per unit frequency band, and represents the spectrum density of the mean power of process $\xi^{(k)}(t)$.

The spectrum density $G_k(\omega)$ can exist even when a Fourier transformation of $\xi^{(k)}(t)$does not exist, i.e., when $\lim_{T\to\infty} Z_{kT}(i\omega) = \infty$. The introduction, in place of the usual spectrum characteristic $Z_{kT}(i\omega)$, of the spectrum density $G_k(\omega)$ is just that

---

* The property of the spectrum of a function limited in time, expressed by (5.41), is analogous to the property of a function with a limited spectrum, which comprises the content of the well-known theorem of V. A. Kotel'nikov, according to which the function $f(t)$, having a limited spectrum, is completely determined by its discrete values at points situated at a distance of $1/2F$ with respect to one another, where F is the maximum frequency in the spectrum of $f(t)$

generalization of harmonic analysis which makes it possible to expend the usual spectrum concepts to functions which do not satisfy the condition of absolute integrability (i.e., those for which a Fourier integral does not exist). An example of such functions are the $\xi^{(k)}(t)$ realizations, under discussion of a random process.

Let us show that the spectrum (Fourier transformation), of the product of $\xi_T^{(k)}(t)\ \xi_T^{(k)}(t+\tau)$ averaged over time, coincides with $2/T\,|Z_{kT}(i\omega)|^2$.

Truly,

$$2\int_{-\infty}^{\infty}\left[\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\xi_T^{(k)}(t)\,\xi_T^{(k)}(t+\tau)\,dt\right]e^{-i\omega\tau}\,d\tau=$$

$$=\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\xi_T^{(k)}(t)\,e^{i\omega t}dt\int_{-\infty}^{\infty}\xi_T^{(k)}(t+\tau)\,e^{-i\omega(t+\tau)}d\tau=$$

$$=\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\xi_T^{(k)}(t)\,e^{i\omega t}dt\int_{-\frac{T}{2}}^{\frac{T}{2}}\xi_T^{(k)}(t_1)\,e^{-i\omega t_1}dt_1,$$

from which, taking into account (5.39), we obtain

$$\int_{-\infty}^{\infty}\left[\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\xi_T^{(k)}(t)\,\xi_T^{(k)}(t+\tau)\,dt\right]e^{-i\omega\tau}\,d\tau=$$

$$=\frac{2}{T}Z_{kT}(-i\omega)\,Z_{kT}(i\omega)=\frac{2}{T}|Z_{kT}(i\omega)|^2.$$

Passing to the limit when $T\to\infty$, we find the spectrum density $G_k(\omega)$ of the mean power of realization $\xi^{(k)}(t)$

$$G_k(\omega)=\lim_{T\to\infty}\frac{2}{T}|Z_{kT}(i\omega)|^2=$$
$$=\int_{-\infty}^{\infty}\left\{\lim_{T\to\infty}\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}\xi_T^{(k)}(t)\,\xi_T^{(k)}(t+\tau)\,dt\right\}e^{-i\omega\tau}\,d\tau. \qquad (5.42)$$

For a stationary random process, the limit enclosed by the curved brackets does not depend on the choice of the realization $\xi^{(k)}(t)$ of a random process and is equal to the correlation function of the process (cf. Section 3).

Thus, for a stationary random process, the spectrum density $G_k(\omega)$ of the mean power of the process $\xi^{(k)}(t)$ is the same for any realization of the random process and is equal to the spectrum $F(\omega)$ of the correlation function $B(\tau)$

$$G_k(\omega)=\lim_{T\to\infty}\frac{2}{T}\,|Z_{kT}(i\omega)|^2=2\int_{-\infty}^{\infty}B(\tau)\,e^{-i\omega\tau}d\tau=F(\omega). \qquad (5.42')$$

This spectrum density of the mean power of any realization of a process, coinciding with the spectrum $F(\omega)$ (Fourier transformation) of correlation function $B(\tau)$, is called the power spectrum of a stationary random process. This spectrum provides only an average power picture, the distribution of the power of a process among the frequencies of the elementary harmonic components, but does not account for the instantaneous phases of these components.

As has been shown by A. Ya. Khinchin, the correlation function of a stationary random process may always be presented in the form of an integral

$$B(\tau)=\frac{1}{4\pi}\int_{-\infty}^{\infty}F(\omega)\,e^{i\omega\tau}d\omega=\frac{1}{2\pi}\int_{0}^{\infty}F(\omega)\cos\omega\tau\,d\omega, \qquad (5.43)$$

where $F(\omega)$ is a limited function $(F(\omega)\geqslant 0)$. Conversely, every function $B(\tau)$, which may be presented in the form of (5.43), is the correlation function of some stationary random process. The function $F(\omega)$ is the direct Fourier transformation of the correlation function:

$$F(\omega)=2\int_{-\infty}^{\infty}B(\tau)\,e^{-i\omega\tau}d\tau=4\int_{0}^{\infty}B(\tau)\cos\omega\tau\,d\tau. \qquad (5.44)$$

This function has, as has been shown above, a clear physical sense: it represents the spectrum density of the mean power of the process.

Let us remark that in formula (5.43), with employment of the Fourier transformation in an exponential form, the concept of the spectrum distribution of the mean power of a process was extended over all real frequencies from $\omega=-\infty$ to $\omega=+\infty$. Only the positive frequencies of from $\omega=0$ to $\omega=+\infty$ make physical sense. To make possible the employment of the exponential form of the Fourier integral, which simplifies

its computation, each spectrum component is broken down into the two equally intensive components $\frac{1}{2}F(\omega)$ and $\frac{1}{2}F(-\omega)$ in such a manner that the total power spectrum $F(\omega)$, distributed over the negative frequencies, becomes an even function of the frequency (which can also be seen from (5.44), if it is remembered that the correlation function is always even). It should, however, always be remembered that the property of the evenness of a power spectrum is preserved only if the origin of the coordinates lies at the point of $\omega = 0$, and may be violated if the origin is displaced to another point.

Since a Fourier transformation is possible only under the condition of the absolute integratibility of a function, therefore formulas (5.43) and (5.44) are valid under the condition that

$$\int_{-\infty}^{\infty}|B(\tau)|\,d\tau \leqslant M, \quad \int_{-\infty}^{\infty}F(\omega)\,d\omega \leqslant N, \tag{5.45}$$

where M and N are constants.

Condition (5.45) restricts the applicability of formulas (5.43) and (5.44) only for stationary processes, the mean value of which is equal to zero (no constant component), and which also have no periodic components. In the work of Khinchin [2] the property in question of stationary random processes is formulated in a more general form, so that the limitations of (5.45) are insignificant. Below it will be shown how formulas (5.43) and (5.44) are generalized for a stationary process, containing both a constant and periodic component.

From (5.43) when $\tau = 0$ we find that the mean power of a stationary process is equal to

$$B(0) = \frac{1}{2\pi}\int_{0}^{\infty}F(\omega)\,d\omega, \tag{5.46}$$

i.e., is equal to the area of its power spectrum. The spectral density of the mean power when $\omega = 0$, as can be seen from (5.44), is equal to

$$F(0) = 2\int_{-\infty}^{\infty}B(\tau)\,d\tau, \tag{5.47}$$

i.e., to double the area under the curve of the correlation function. The ratio $\frac{F(0)}{4\,B(0)}$ may, in accordance with (5.32), be taken as a measure of the time of correlation.

The correlation function $B(\tau)$ and the power spectrum $F(\omega)$ of a stationary random process, as a pair of Fourier transformations, possess all the properties intrinsic to this transformation. Specifically, the "wider" is the spectrum $F(\omega)$, the "narrower" is the correlation function $B(\tau)$, and vice versa. In other words, the product of the correlation time τ_0 and bandwidth Δ of the power spectrum is a constant.

For a nonstationary random process, the spectrum density $G_{(k)}(\omega)$ for various realizations $\xi^{(k)}(t)$ of a random process varies. Therefore for a nonstationary random process the power spectrum of the process is called the average of an aggregate of $G_k(\omega)$, i.e., the quantity

$$m_1\{G_k(\omega)\} = m_1\left\{\lim_{T\to\infty}\frac{2}{T}\,|Z_{kT}(i\omega)|^2\right\} = \\ = \lim_{T\to\infty}\frac{2}{T}\,m_1\{|Z_{kT}(i\omega)|^2\}. \tag{5.48}$$

It follows from (5.48) that the power spectrum of a nonstationary random process coincides with the spectrum (Fourier transformation) averaged for the time of the correlation function of this process

$$F(\omega) = \int_{-\infty}^{\infty}\left\{\lim_{T\to\infty}\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} B(\tau,t)\,dt\right\} e^{-i\omega\tau}\,d\tau = \\ = \lim_{T\to\infty}\frac{2}{T}\,m_1\{|Z_{kT}(i\omega)|^2\}. \tag{5.49}$$

Formula (5.49) is a generalization of (5.42') for nonstationary random processes.

From what has been stated above, it follows that two ways exist for computing the power spectrum of a random process:

a) direct, proceeding from the observation of a single realization $\xi^{(k)}(t)$ and then finding the limit, $\lim_{T\to\infty} 2/T\,\left|Z_{kT}(i\omega)\right|^2$, or the average of an aggregate of this limit for a nonstationary random process;

b) finding the Fourier transformation of a correlation function for a stationary

random process, or a correlation function averaged over time — for a nonstationary one.

7. Power Spectrum of Generalized Telegraph Signal

Let us, as an example, compute the correlation function and power spectrum for the generalized signal $\xi(t)$*, one of the realizations of which appears in Figure 2 (p. 34). It can be shown that the probability of exactly k sign changes taking place in the signal, during the time t, is determined by the formula of Poisson (1.54) (cf. Section 9, Ch. 1)

$$P(k)=\frac{(\nu t)^k}{k!}e^{-\nu t}$$

upon the fulfillment of the following condition: no matter how many sign changes take place in the time period (0, t), the probability of one sign change over the interval of $(t, t+\Delta t)$ is equal to $\nu\Delta t$, and the probability of more than one sign change decreases more rapidly than Δt. Here ν is the constant average number of sign changes in a unit of time, and νt is the average number of sign changes during the time t. In the general case, the mean density of the points of intersection of the curve with the abscissa (cf. Fig. 2) may be a function of time.

Designating this function by $\nu(t)$, we find the average number of sign changes over the time interval of from t_1 to t_2.

$$\lambda=\int_{t_1}^{t_2}\nu(t)\,dt. \tag{5.50}$$

Since $\nu(t)\geqslant 0$, therefore $\lambda>0$.

The probability that, over the time interval from t_1 to t_2 exactly k sign changes of the signal will take place, will be equal to

$$P(k)=\frac{\lambda^k e^{-\lambda}}{k!}. \tag{5.51}$$

* The generalized telegraph signal is a special form of the so-called Poisson random process, a detailed investigation of which appears in [5] .

Formula (1.54) is a special case of (5.51), when $\nu(t)$ is a positive constant.

To compute the correlation function of a generalized telegraph signal it is necessary, according to definition (5.16), to obtain an average of the aggregate realizations of the product of $\xi(t_1)\ \xi(t_2)$. This product may be equal either to h^2, or $-h^2$ (cf. Fig. 2) depending on whether $\xi(t_1) = \xi(t_2)$ or $\xi(t_1) = -\xi(t_2)$. The equality $\xi(t_1) = \xi(t_2)$ signifies that, over the interval from t_1 to t_2, an even number of sign changes occurred, i.e., there occurred one of the mutually incompatible events: $k = 0$, or $k = 2$, or $k = 4, \ldots$

Then, employing (5.51), we find by the rule of addition

$$P\{\xi(t_1)=\xi(t_2)\}=\sum_{k=0}^{\infty}P(2k)=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!}=e^{-\lambda}\operatorname{ch}\lambda. \tag{5.52}$$

Analogously, the equality $\xi(t_1) = -\xi(t_2)$ signifies that, over the interval of from t_1 to t_2, there occurred an odd number of sign changes and

$$\begin{gathered}P\{\xi(t_1)=-\xi(t_2)\}=\sum_{k=1}^{\infty}P(2k-1)=\\=e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{2k-1}}{(2k-1)!}=e^{-\lambda}\operatorname{sh}\lambda.\end{gathered} \tag{5.53}$$

Now it is not difficult to obtain an average of the product, i.e., the correlation function

$$\begin{gathered}B(t_1,\ t_2)=m_1\{\xi(t_1)\xi(t_2)\}=h^2P\{\xi(t_1)=\xi(t_2)\}-\\-h^2P\{\xi(t_1)=-\xi(t_2)\}.\end{gathered} \tag{5.54}$$

Substituting (5.52) and (5.53) into (5.54), we obtain

$$B(t_1,\ t_2)=h^2e^{-\lambda}(\operatorname{ch}\lambda-\operatorname{sh}\lambda)=h^2e^{-2\lambda},$$

or, taking into account (5.50),

$$B(t_1,\ t_2)=h^2e^{-2\int_{t_1}^{t_2}\nu(t)dt} \tag{5.55}$$

If the generalized telegraph signal is stationary, then $\nu(t) = \nu_0$ = const, and the correlation function depends only of the difference $\tau = t_2 - t$

$$B(\tau)=h^2e^{-2\nu_0|\tau|}. \tag{5.56}$$

(By virtue of the evenness of $B(\tau)$, the absolute value of τ has been used in the exponent.)

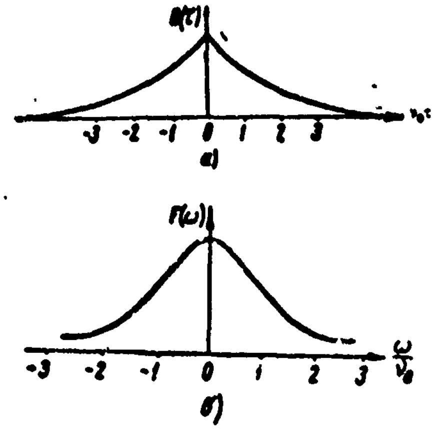


Fig. 41. a) Correlation function and b) power spectrum of a generalized telegraph signal.

A graph of this correlation function is shown in Figure 41 a. Employing the Khinchin theorem [formula (5.44)], it is now not difficult also to find the power spectrum of a stationary telegraph signal (Fig. 41 b).

$$F(\omega) = 4h^2 \int_0^{\infty} e^{-2\nu_0\tau} \cos \omega\tau d\tau = \frac{8h^2\nu_0}{\omega^2 + 4\nu_0^2}. \tag{5.57}$$

Let us now examine an example of a nonstationary process. In this case, the correlation function depends on two variables, and a supplementary averaging is required of one of them:

$$B(\tau) = h^2 \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} e^{-2\lambda(t,\tau)} dt.$$

Let, for instance, the mean density of the intersection points be a periodic function of time

$$\nu(t) = \nu_0 + \nu_1 \cos \omega_0 t. \tag{5.58}$$

From the condition $\nu(t) \geqslant 0$ there follows $\nu_1 \leqslant \nu_0$. Substituting (5.58) into (5.50), we find

$$\lambda(t,\tau)=\int_t^{t+\tau}(\nu_0+\nu_1\cos\omega_0 t)\,dt=\nu_0\tau+$$
$$+\frac{\nu_1}{\omega_0}[\sin\omega_0 t(\cos\omega_0\tau-1)+\cos\omega_0 t\sin\omega_0\tau]=\nu_0\tau+$$
$$+\frac{\nu_1}{\omega_0}2\sin\frac{\omega_0\tau}{2}\cos(\omega_0 t+\varphi).$$

Since $\lambda(t,\tau)$ is a periodic function with a period of $T = \frac{2\pi}{\omega_0}$, it is sufficient to average the correlation function over the time T

$$B(\tau)=\frac{h^2 e^{-2\nu_0\tau}}{T}\int_0^T e^{-\frac{4\nu_1}{\omega_0}\sin\frac{\omega_0\tau}{2}\cos(\omega_0 t+\varphi)}dt$$

or (cf. (3.37))

$$B(\tau)=h^2 e^{-2\nu_0|\tau|}I_0\left(\frac{4\nu_1}{\omega_0}\sin\frac{\omega_0\tau}{2}\right), \tag{5.59}$$

where $I_0(x)$ is a Bessel function of an imaginary, zero-order argument. As before, an absolute value of τ is taken in the exponent by reason of the evenness of the correlation function. When $\nu_1 = 0$, formula (5.59) turns into (5.56).

To determine the power spectrum it is useful first to expand the Bessel function into a Fourier series

$$I_0\left(\frac{4\nu_1}{\omega_0}\sin\frac{\omega_0\tau}{2}\right)=\sum_{k=-\infty}^{\infty}c_k e^{ik\omega_0\tau},$$

whose coefficients are

$$c_k=c_{-k}=\sum_{n=k}^{\infty}(-1)^k\left(\frac{\nu_1}{\omega_0}\right)^{2n}\frac{1}{(n!)^2}\frac{(2n)!}{(n-k)!\,(n+k)!},\quad k>0.$$

Now, subjecting (5.59) to a Fourier transformation, we find the power spectrum, averaged over time, of the generalized telegraph signal under examination

$$F(\omega)=2h^2\sum_{k=-\infty}^{\infty}c_k\int_{-\infty}^{\infty}e^{-2\nu_0|\tau|}e^{i(k\omega_0-\omega)\tau}d\tau,$$

or

$$F(\omega)=\sum_{k=-\infty}^{\infty}F_k(\omega), \tag{5.60}$$

where

$$F_k(\omega)=\frac{8h^2c_k\nu_0}{(\omega-k\omega_0)^2+4\nu_0^2}.$$

From a comparison of (5.60) with (5.57), it follows that in the case at hand, the power spectrum constitutes a linear combination of the same type of curves as in Figure 41 b, whose maxima lie in frequencies of k ω_o.

8. Continuity and Differentiability of a Stationary Random Process

In the majority of engineering applications, the correlation function B(τ) of a stationary random process is a continuous function of this argument τ. It is possible to show that, for continuity with any τ, it is sufficient that a correlation function be continuous at one single point of τ = 0 [1].

The continuity of a correlation function provides also for the continuity of the random stationary process, which signifies that for any moment of time t and for an arbitrarily small $\varepsilon > 0$,

$$\lim_{T\to 0} P\{|\xi(t+T)-\xi(t)|>\varepsilon\}=0.$$

The random process ξ(t) is differentiable at the point t, if there exists such a random variable ξ'(t), called the derivative of process ξ(t) with respect to point t, that

$$\lim_{T\to 0} P\left\{\left|\frac{\xi(t+T)-\xi(t)}{T}-\xi'(t)\right|>\varepsilon\right\}=0.$$

Thus, the statistical characteristics of the derivative ξ'(t) may be obtained by a limit transition from the statistical characteristics of the difference of ξ(t + T) — ξ(t). In certain problems, calculation of the statistical characteristics of the difference ξ(t + T) — ξ(t) may be independently significant. We shall restrict ourselves to determining the one-dimensional distribution function of this difference, and its correlation function and power spectrum.

If the two-dimensional distribution function $w_2(x_1, x_2, T)$ of the process ξ (t) is known, then employing (3.24'), we calculate the one-dimensional distribution

function of the difference of the random variables

$$W_1(x, T) = \int_{-\infty}^{\infty} w_2(x_1, x + x_1, T)\, dx_1. \tag{5.61}$$

It is not difficult to find the correlation function $B_T(\tau)$ of this difference, if the correlation function $B(\tau)$ of the stationary random process $\xi(t)$ is known

$$\begin{aligned} B_T(\tau) &= m_1\{[\xi(t+T) - \xi(t)][\xi(t+T+\tau) - \xi(t+\tau)]\} = \\ &= m_1\{\xi(t+T)\xi(t+T+\tau)\} - m_1\{\xi(t)\xi(t+T+\tau)\} - \\ &\quad - m_1\{\xi(t+T)\xi(t+\tau)\} + m_1\{\xi(t)\xi(t+\tau)\}, \end{aligned}$$

wherefrom

$$B_T(\tau) = 2B(\tau) - B(\tau+T) - B(\tau-T). \tag{5.62}$$

The power spectrum $F_T(\omega)$ of the difference is linked by a simple relationship to the power spectrum $F(\omega)$ of the process $\xi(t)$

$$\begin{aligned} F_T(\omega) &= 2\int_{-\infty}^{\infty} B_T(\tau) e^{-i\omega\tau} d\tau = 4\int_{-\infty}^{\infty} B(\tau) e^{-i\omega\tau} d\tau - \\ &- 2\int_{-\infty}^{\infty} B(\tau+T) e^{-i\omega\tau} d\tau - 2\int_{-\infty}^{\infty} B(\tau-T) e^{-i\omega\tau} d\tau = \\ &= 2\int_{-\infty}^{\infty} B(\tau)[2e^{-i\omega\tau} - e^{-i\omega(\tau-T)} - e^{-i\omega(\tau+T)}] d\tau = \\ &= 2\cdot 2(1 - \cos\omega T)\int_{-\infty}^{\infty} B(\tau) e^{-i\omega\tau} d\tau, \end{aligned}$$

or

$$F_T(\omega) = 4\sin^2\frac{\omega T}{2} \cdot F(\omega). \tag{5.63}$$

From (5.61) - (5.63) it is possible to find the corresponding statistical characteristics of the derivative $\xi'(t)$.

In view of (3.12) we find that the distribution function of $\frac{\xi(t+T) - \xi(t)}{T}$ is equal to $TW_1(Tx_1, T)$, and the distribution function $W_1^{(1)}(x)$ of the derivative $\xi'(t)$ is obtained by taking the limit

$$W_1^{(1)}(x) = \lim_{T\to 0} TW_1(Tx_1, T).$$

Since the correlation function of $\frac{\xi(t+T)\cdot\xi(t)}{T}$ is equal to $\frac{1}{T^2} B_T(\tau)$ — therefore from (5.62) expanding the right hand side into a Taylor series, we obtain for the correlation function $B_1(\tau)$ of the derivative $\xi'(t)$ the following expression

$$B_1(\tau)=\lim_{T\to 0}\frac{1}{T^2}B_T(\tau)=\lim_{T\to 0}\{-B''(\tau)+O(T^2)\}=-B''(\tau). \quad (5.64)$$

Thus the correlation function of the derivative $\xi'(t)$ is equal to the second derivative of the correlation function of $\xi(t)$, with the opposite sign.

From (5.53) by means of taking the limit, we also find the power spectrum $F_1(\omega)$ of the derivative

$$F_1(\omega)=\lim_{T\to 0}\frac{1}{T^2}4\sin^2\frac{\omega T}{2}F(\omega)=\omega^2F(\omega). \quad (5.65)$$

Formula (5.65) follows also from (5.64) and (5.43).

The dispersion of the derivative (with a zero mean value) is equal to

$$B_1(0)=-B''(0)=\frac{1}{2\pi}\int_0^\infty\omega^2F(\omega)\,d\omega. \quad (5.64')$$

Since when $\tau = 0$ the correlation function $B(\tau)$ always attains a maximum, $B''(0)<0$.

Let us now determine the correlation function of the differentiable stationary process $\xi(t)$ and of its derivative $\xi'(t)$.

Since

$$m_1\{\xi(t)[\xi(t+T+\tau)-\xi(t+\tau)]\}=B(T+\tau)-B(\tau),$$

the desired mutual correlation function is

$$m_1\{\xi(t)\xi'(t+\tau)\}=\lim_{T\to 0}m_1\left\{\xi(t)\frac{\xi(t+\tau+T)-\xi(t+\tau)}{T}\right\}=$$
$$=\lim_{T\to 0}\frac{B(T+\tau)-B(\tau)}{T}=B'(\tau),$$

i.e.,

$$m_1\{\xi(t)\xi'(t+\tau)\}=B'(\tau). \quad (5.66)$$

Considering (5.43), we also find

$$m_1\{\xi(t)\xi'(t+\tau)\}=-\frac{1}{2\pi}\int_0^\infty\omega F(\omega)\sin\omega\tau\,d\omega \quad (5.66')$$

When $\tau = 0$, from (5.66') there follows

$$m_1\{\xi(t)\xi'(t)\}=B'(0)=0. \quad (5.67)$$

Thus the mutual correlation function of a stationary random process and of its derivative in coincident time instants is always equal to zero, i.e., a random function and its derivative are, in coincident time instants, not correlated. From this it follows that the joint distribution function of $\xi(t)$ and $\xi'(t)$ is equal simply to the product of their one-dimensional distribution functions.

$$w_2(x, x') = w_1(x) \cdot w_1^{(1)}(x'). \qquad (5.67')$$

There may also be examined the n-th derivative $\xi^{(k)}(t)$ of the random process $\xi(t)$, for the existence of which it is sufficient that there exist a continuous derivative of the 2n-th order of the correlation function of the process $\xi(t)$. The correlation function of the n-th derivative of the process is equal to

$$B_n(\tau) = (-1)^n B^{(2n)}(\tau),$$

and its power spectrum to

$$F_n(\omega) = \omega^{2n} F(\omega).$$

9. Random Processes with a Discrete Spectrum.

If the conditions of (5.45) are fulfilled, then the power spectrum $F(\omega)$ of a random process is a continuous, non-negative function of the frequency ω.

Stationary random processes with a continuous spectrum are not, however, the only such processes. There exist stationary (in the wide sense) processes, the correlation functions of which with an unlimited increase of τ tend toward a constant, or are periodic functions of τ. This, of course, violates the first condition of (5.45) and the power spectrum ceases to be a continuous function of the frequency.

Let us, for instance, examine the random process

$$\xi(t) = \xi_1 \cos \omega_1 t + \xi_2 \sin \omega_1 t, \qquad (5.68)$$

where ξ_1 and ξ_2 are random variables not dependent on t, and ω_1 is a constant.

It can be said that the random function $\xi(t)$ represents "harmonic vibrations" with a random amplitude of $A = \sqrt{\xi_1^2 + \xi_2^2}$ and a phase of $\varphi = \text{arc tg} \frac{\xi_2}{\xi_1}$, the

distributions of which do not depend on time.

Random process (5.68) will not always be stationary. For its stationarity (in the wide sense) it is necessary that the following conditions be fulfilled. First, the mean value of the process may not depend on time, inasmuch as

$$m_1\{\xi(t)\}=m_1\{\xi_1\}\cos\omega_1 t+m_1\{\xi_2\}\sin\omega_1 t,$$

this condition may only be fulfilled when

$$m_1\{\xi_1\}=m_1\{\xi_2\}=0, \tag{5.69}$$

i.e., when the mean values of the random variables ξ_1 and ξ_2 are equal to zero. In the second place, the correlation function of the process may depend only on the one parameter τ, inasmuch as

$$\begin{aligned}B(t,\tau)=m_1\{\xi(t)\xi(t+\tau)\}=m_1\{\xi_1^2\cos\omega_1 t\cos\omega_1(t+\tau)+\\+\xi_2^2\sin\omega_1 t\sin\omega_1(t+\tau)+\xi_1\xi_2\cos\omega_1 t\sin\omega_1(t+\tau)+\\+\xi_1\xi_2\sin\omega_1 t\cos\omega_1(t+\tau)\}=m_1\{\xi_1^2\}\cos\omega_1 t\cos\omega_1(t+\tau)+\\+m_1\{\xi_2^2\}\sin\omega_1 t\sin\omega_1(t+\tau)+m_1\{\xi_1\xi_2\}\sin\omega_1(2t+\tau).\end{aligned}$$

$B(t,\tau)$ will not depend on t, if

$$m_1\{\xi_1^2\}=m_1\{\xi_2^2\}=\frac{\sigma^2}{2} \tag{5.70}$$

and

$$m_1\{\xi_1\cdot\xi_2\}=0. \tag{5.71}$$

The latter is fulfilled if ξ_1 and ξ_2 are independent.

When conditions (5.69) - (5.71) are fulfilled, a harmonic vibration with random amplitude and phase will be stationary, and its correlation function will be equal to

$$B(\tau)=\frac{\sigma^2}{2}\cos\omega_1\tau. \tag{5.72}$$

In this case the amount of dispersion of the random amplitude will be

$$M_2\{A\}=m_1\{\xi_1^2+\xi_2^2\}=m_1\{\xi_1^2\}+m_1\{\xi_2^2\}=\sigma^2.$$

As can be seen from (5.72), the correlation function of vibration with a random

amplitude and phase is proportional to the dispersion of the amplitude, but does not depend on any statistical characteristic of the phase.

Although correlation function (5.72) does not satisfy inequality (5.45), the concept of the power spectrum can nevertheless be expanded also to the case at hand, with the employment of the delta-function (cf. Appendix IV). Then the Fourier transformation of correlation function (5.72), i.e., the power spectrum of vibration with a random amplitude and phase, may be represented in the form of

$$F(\omega)=\frac{\pi a^2}{2}[\delta(\omega+\omega_1)+\delta(\omega-\omega_1)]. \qquad (5.73)$$

This spectrum consists of two discrete lines of unlimited intensity at the frequencies of $\pm\,\omega_1$.

A further generalization is a random process which is the superposition of n elementary stationary random processes of the (5.68) type:

$$\xi(t)=\sum_{k=1}^{n}(\xi_k\cos\omega_k t+\eta_k\sin\omega_k t). \qquad (5.74)$$

The conditions of stationarity (in the broad sense) of this process are analogous to (5.69) - (5.71). The correlation function of process (5.74) takes the following form when the conditions of stationarity are fulfilled:

$$B(\tau)=\sum_{k=1}^{n}\frac{\sigma_k^2}{2}\cos\omega_k\tau, \qquad (5.75)$$

where

$$\frac{\sigma_k^2}{2}=m_1\{\xi_k^2\}=m_1\{\eta_k^2\}.$$

The power spectrum of stationary process (5.74), i.e., the Fourier transformation of correlation function (5.75), is the sum of the delta-functions at the discret frequencies of $\pm\,\omega_K$

$$F(\omega)=\sum_{k=1}^{n}\frac{\pi\sigma_k^2}{2}[\delta(\omega+\omega_k)+\delta(\omega-\omega_k)]. \qquad (5.76)$$

A power spectrum of the type of (5.76) is called discrete, and the stationary random process to which it corresponds is called a process with a discrete spectrum.

It is evident from a comparison of (5.75) and (5.22), that the sum of S(t) harmonic vibrations with frequencies of ω_K and amplitudes of c_k has the same autocorrelation function, and consequently the same power spectrum, as a stationary random process representing the sum of the same frequencies with random amplitudes and phases, the dispersions of the random amplitudes coinciding with c_k^2. This comparison emphasizes once again the fact that the correlation function provides only an average power concept of a process, without taking momentary values into account, as a result of which it could turn out that two processes which are different in principle, have the same correlation functions and power spectra.

A stationary random process with a discrete spectrum of the (5.74) type

$$\xi(t)=\sum_{k=1}^{\infty}\left(\xi_k\cos\frac{2\pi k}{T}t+\eta_k\sin\frac{2\pi k}{T}t\right),$$

where ξ_K and η_K are random variables subject to the conditions of

$$m_1\{\xi_k\}=m_1\{\eta_k\}=0,$$

$$m_1\{\xi_k\xi_n\}=m_1\{\eta_k\eta_n\}=\begin{cases}\sigma_k^2/2\,(k=n),\\ 0\quad (k\neq n),\end{cases}$$

$$m_1\{\xi_k\eta_n\}=0,$$

is sometimes regarded as an approximation of a stationary random process with a continuous spectrum. This approximation signifies the following. The realization of a random process with a continuous spectrum is observed during the time NT, where N is very large. Then the total period of observation is broken down into N equal parts, each with a duration of T. For each of the indicated time segments T, the realization of the random process may be resolved into a Fourier series

$$\xi_T^{(r)}(t)=\sum_{k=1}^{\infty}a_k^{(r)}\cos\frac{2\pi k}{T}t+b_k^{(r)}\sin\frac{2\pi k}{T}t,\quad r=1,2,\ldots N.$$

The coefficients a_k and b_k change with a transfer from one segment to another. The aggregates of the values of these coefficients are the random variables ξ_K and η_K in (5.74). It is clear, that the larger is T, the better is the indicated approximation.

This approximation may also be represented in another form

$$\xi(t)=\sum_{k=1}^{\infty}\zeta_k\cos\left(\frac{2\pi k}{T}t+\varphi_k\right).$$

With such a representation, $\xi(t)$ is regarded as a sum of harmonic vibrations with random amplitudes and phases.

10. Narrow-band and Wide-band Processes.

A case of practical importance is when the power spectrum $F(\omega)$ of a random process is primarily concentrated in a relatively narrow frequency band about a certain fixed high frequency ω_0, where the spectral density is at its maximum (Fig. 42). Then the correlation function $B(\tau)$ of such a random process may be represented in the form of

$$B(\tau)=\frac{1}{2\pi}\int_0^\infty F(\omega)\cos\omega\tau\, d\omega=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\omega_0} F(\omega_0-\omega)\cos(\omega_0-\omega)\tau\, d\omega=$$

$$=\left[\frac{1}{2\pi}\int_{-\infty}^{\omega_0} F(\omega_0-\omega)\cos\omega\tau\, d\omega\right]\cos\omega_0\tau+$$

$$+\left[\frac{1}{2\pi}\int_{-\infty}^{\omega_0} F(\omega_0-\omega)\sin\omega\tau\, d\omega\right]\sin\omega_0\tau. \tag{5.77}$$

Since, according to the assumption, the spectrum band is relatively narrow in comparison to ω_0, the upper limits of integration in (5.77) may without appreciable error be extended to infinity.

Designating

$$F^*(\omega)=F(\omega_0-\omega), \tag{5.78}$$

$$a(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty} F^*(\omega)\cos\omega\tau\, d\omega, \tag{5.79}$$

$$b(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty} F^*(\omega)\sin\omega\tau\, d\omega, \tag{5.80}$$

we obtain

$$B(\tau)=a(\tau)\cos\omega_0\tau+b(\tau)\sin\omega_0\tau. \tag{5.81}$$

In virtue of the fact that power spectrum $F(\omega)$ is concentrated in a narrow frequency band about $\omega=\omega_0$, the spectrum $F(\omega_0-\omega)=F^*(\omega)$ lies in a low-frequenc

range. Then, as can be seen from (5.79) and (5.80), $a(\tau)$ and $b(\tau)$ will be slowly changing functions of τ. Here, if the bandwidth of the low-frequency spectrum is of the order of Δ, the correlation function will have a "width" on the order of $1/\Delta$. If $F(\omega)$ may be considered symmetrical with respect to the central frequency ω_0, then $b(\tau) = 0$, and from (5.81) we obtain

$$B(\tau)=a(\tau)\cos\omega_0\tau=\left[\frac{1}{\pi}\int_0^\infty F^*(\omega)\cos\omega\tau\,d\omega\right]\cos\omega_0\tau. \tag{5.82}$$

Consequently, the correlation function of a narrow-band process, whose spectrum is symmetrical about a high frequency ω_0, is equal to the correlation function $a(\tau)$ multiplied by $\cos\omega_0\tau$, which corresponds to spectrum $F^*(\omega)$ obtained from the original one through displacement by the amount of ω_0 into the low-frequency range.

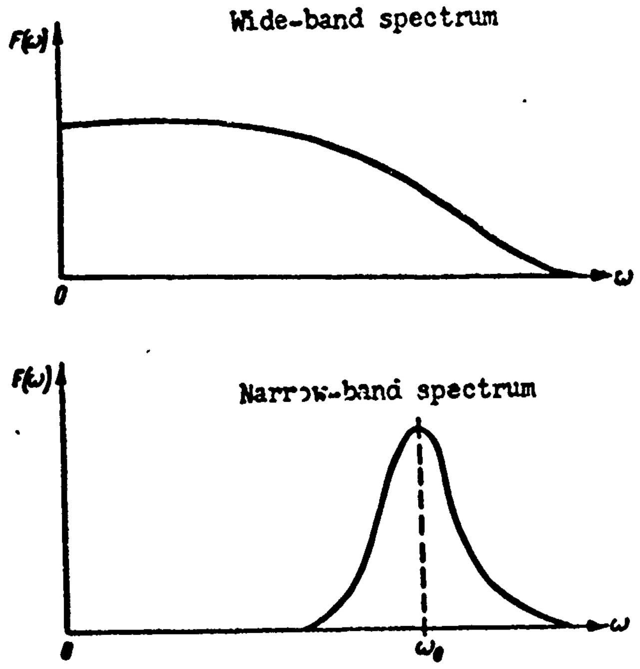


Fig. 42. Narrow-band and wide-band power spectra.

In contrast to the narrow-band spectrum examined above, let us now examine the power spectrum of a random process with a very wide band (Fig. 42). Let the spectrum density $F(\omega)$ of the mean power of the random process maintain a constant value up to very high frequencies. The correlation function $B(\tau)$ of such a process will differ

from zero only in a very small interval of the values of its argument about the origin of the coordinates, i.e., for small instances of τ. The power spectrum

$$F(\omega)=F_0=\text{const}, \qquad (5.83)$$

uniform at all frequencies, is a useful mathematical idealization of the spectra of the indicated type.

A random process with a uniform spectrum at all frequencies is called "white noise"*. The correlation function of white noise is equal to

$$B(\tau)=\frac{1}{4\pi}\int_{-\infty}^{\infty}F_0e^{i\omega\tau}d\omega=\frac{F_0}{2}\delta(\tau), \qquad (5.84)$$

i.e., is a delta-function at the origin of its coordinates (cf. appendix IV).

The correlation coefficient for white noise is

$$R(\tau)=\begin{cases}1 & \tau=0,\\ 0 & \tau\neq 0.\end{cases} \qquad (5.85)$$

Thus the white noise $\xi(t)$ is characterized by the fact that the values of $\xi(t)$ at any two time instants (no matter how close together) are uncorrelated. For this reason white noise is sometimes also called an absolutely random process.

It should be emphasized, that the white noise concept defines only the spectrum picture of a random process and leaves completely open the question of the laws of distribution. Since the power spectrum does not uniquely define the laws of distribution, white noise may be called random processes which have a uniform power spectrum and various laws of distribution.

Thus, white noises are an example of those random processes whose correlation functions and power spectra are the same, and whose distribution laws may differ from one another.

If a sequence of base pulses is employed as an approximation to the stationary random process, the case where the amplitudes of these pulses are mutually independent will correspond to white noise. For this reason various types of white noise

* In analogy to white light, which has a continuous and approximately homogenous spectrum within the limits of its visible portion.

must be fully characterized by only a one-dimensional distribution function. It is, for instance, possible to speak of white noise with a normal distribution law or of white noise with a Rayleigh distribution law.

White noise is an idealization never realized under actual conditions since, in the first place, sufficiently close values of a random function are practically always dependent and, secondly, real processes have finite power, while for white noise the full power of the process is infinite. However, as a result of the limited nature of the pass bands of radio equipment, such an idealization, while considerably simplifying the mathematical analysis of the processes, introduces no errors of any significance*.

11. The Normal Random Process.

Some general statistical characteristics of stationary random processes have been cited above, and their properties have been indicated. The special type of these processes most frequently encountered in radio-engineering applications is the so-called normal random process.

The random process $\xi(t)$ is called normal (gaussian) if the distribution functions of any n-th order for the aggregate of random variables $\xi_k = \xi(t_k)$ (k=1, 2,...n) are normal, i.e., are determined by the formulas (cf. (2.40))

$$w_n(x_1, \ldots x_n, t_1, \ldots t_n) = \frac{1}{\sqrt{(2\pi)^n D \sigma_1 \sigma_2 \ldots \sigma_n}} \times$$

$$\times e^{-\frac{1}{2D} \sum_{l=1}^{n} \sum_{k=1}^{n} D_{lk} \frac{x_l - a_l}{\sigma_l} \cdot \frac{x_k - a_k}{\sigma_k}}, \tag{5.86}$$

where

$$a_k = m_1 \{\xi(t_k)\},$$
$$\sigma_k^2 = m_1 \{[\xi(t_k) - a_k]^2\}. \tag{5.87}$$

$$D = \begin{vmatrix} R_{11} & R_{12} \ldots R_{1n} \\ R_{21} & R_{22} \ldots R_{2n} \\ R_{n1} & R_{n2} \ldots R_{nn} \end{vmatrix}. \tag{5.88}$$

(5.89)

* Strictly speaking, the concept of the value of the random function $\xi(t)$, at the point t, itself ceases to make sense for white noise and it is possible to speak only of the results of the passage of white noise through linear systems, as will be done in the analysis of the following chapter (cf. also [19]).

The quantity D_{ik} is an algebraic co-factor in the determinant D of the element R_{ik}, which is the correlation coefficient of the random variables $\xi(t_i)$ and $\xi(t_k)$.

$$R_{ik}=\frac{m_1\{[\xi(t_i)-a_i][\xi(t_k)-a_k]\}}{\sigma_i\sigma_k}=\frac{B(t_i,t_k)}{\sigma_i\sigma_k}=R(t_i,t_k). \quad (5.90)$$

If all the mean values a_k and the dispersions σ_k are constant (i.e., do not depend on the index k),

$$a_k=a,\quad \sigma_k=\sigma, \quad (5.91)$$

and the correlation function $B(t_i, t_k)$ depends not on the two variables t_i and t_k but only on their difference $\tau_{ik} = t_i - t_k$, then distribution function (5.86) does not change with any displacement of the entire group of points $t_1, \ldots t_n$ by a constant amount along the time axis, i.e., with the fulfillment of the indicated conditions, a normal random process is stationary. Here a distribution function of the n-th order will depend only on the n-1 parameters $\tau_k = |t_i - t_k|$

$$w_n(x_1,\ldots\tau_{n-1})=\frac{1}{\sqrt{(2\pi\sigma^2)^nD}}e^{-\frac{1}{2\sigma^2D}\sum_{i=1}^{n}\sum_{k=1}^{n}D_{ik}(x_i-a)(x_k-a)}, \quad (5.92)$$

where a and σ are constant numbers, while D and D_{ik} are numerical functions of the parameters $\tau_1 \ldots \tau_{n-1}$. These functions are determined by the values of the correlation coefficient $R(\tau) = \frac{B(\tau)}{\sigma^2}$ with the indicated n - 1 values of τ.

Since for a stationary process the correlation coefficient is an even function, therefore in determinant (5.89), for a stationary normal process, $R_{kk} = R(\sigma) = 1$ and $R_{ki} = R_{ik} = R(\tau_k)$, i.e., the elements of the principal diagonal of D are equal to unity, and those of its elements which are symmetrical with respect to this diagonal are equal to each other.

Thus, in order to determine the distribution function of a normal random process of any order, it is sufficient to know only its correlation function. If this correlation function depends on one argument τ (stationarity in the wide sense), the normal process is stationary in the precise sense. Normal stationary processes may differ from each other in the form of their correlation function or of their power

spectrum, given in each concrete problem on the basis of supplementary conditions. If two random normal processes are incoherent i.e., if their mutual correlation function is equal to zero, then they are independent (which, as has been noted above, is not valid in the general case).

The first three distribution functions of a stationary random process with a zero mean value have, in accordance with (2.14), (2.15), and (2.62) the following form:

$$w_1(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}, \tag{5.93}$$

$$w_2(x_1, x_2, \tau)=\frac{1}{2\pi\sigma^2\sqrt{1-R^2(\tau)}}e^{-\frac{x_1^2+x_2^2-2R(\tau)x_1x_2}{2\sigma^2[1-R^2(\tau)]}}, \tag{5.94}$$

$$w_3(x_1, x_2, x_3, \tau_1, \tau_2)=\frac{1}{(2\pi\sigma^2)^{3/2}\sqrt{1-R_1^2-R_2^2-R_3^2+2R_1R_2R_3}}e^{-M}, \tag{5.95}$$

where

$$M=\frac{(1-R_2^2)x_1^2+(1-R_3^2)x_2^2+(1-R_1^2)x_3^2}{2\sigma^2(1-R_1^2-R_2^2-R_3^2+2R_1R_2R_3)}-$$

$$-\frac{2(R_1-R_2R_3)x_1x_2+2(R_2-R_3R_1)x_2x_3+2(R_3-R_2R_1)x_1x_3}{2\sigma^2(1-R_1^2-R_2^2-R_3^2+2R_1R_2R_3)},$$

$$R_1=R_{12}=R_{21}=R(\tau_1),$$
$$R_2=R_{23}=R_{32}=R(\tau_2),$$
$$R_3=R_{31}=R_{13}=R(\tau_1+\tau_2).$$

A normal random process may be a white noise, if the spectrum of the process is uniform at all frequencies. Then the correlation coefficient R (τ) is determined according to formula (5.85) and consequently, $R_{ik} = R_{ki} = 0$ ($k \neq i$), while from (5.89) we find D = 1 and

$$D_{ik}=\begin{cases}1, & i=k\\ 0, & i\neq k\end{cases}.$$

The distribution function of the n-th order in this case, in accordance with (5.92), has the form of

$$w_n(x_1, x_2, \ldots x_n)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k-a)^2},$$

i.e., is the product of n one-dimensional distribution functions, which corresponds to the independence of momentary values of white noise at any two moments in time.

The normal stationary random process with a uniform spectrum embraces a broad class of phenomena observable in radio apparatus.

Employing the central limit theorem, it is possible to show that the so-called internal (fluctuation) noises, caused by the shot effect and by the thermal motion of electrons, have a normal law of distribution, which is also confirmed experimentally. The spectrum of these noises is, according to various data [7], practically uniform to frequencies on the order to 10^{13} cycles, which goes far beyond the limits of that frequency range which is of interest in contemporary radio-engineering problems.

The fact that a normal distribution may be obtained through the addition of a large number of independent vibrations, is employed in experimental practice for the modelling of normally distributed noises through summation of the sinusoidal vibrations of independent generators with appropriately selected amplitudes and frequencies.

Let us note that certain types of internal noise do not have a normal distribution. Among these are, for instance, contact noise, which takes place when current travels along a boundary between two conductors, as well as noise in semiconductors (transistors) [7].

A stationary normal random process with a uniform spectrum in a frequency band of the order of 4 kc may be employed as an approximate model of a modulating voltage corresponding to the transmission of human speech. The validity of such modeling has been confirmed by several experiments [8].

It is all the more permissible to consider as normal the probability distribution of instantaneous values of multichannel communication (with frequency division of channels) which is the sum of a large number of independent speech vibrations.

It is often necessary to study processes which are a sum of the fluctuation noises (normal stationary random process) $\xi(t)$ and the determined signal $S(t)$.

Bearing in mind (5.8), we find that the distribution functions of the sum $\xi(t)$, $S(t)$ are also normal and are obtained from (5.92), if in the exponent of each of the binomials, $x_k - a$ is replaced by $x_k - a - S(t)$. Therefore the sum of signal and noise at hand is a normal stationary process which, however, is not stationary in the strict sense.

12. Derivative of the Normal Random Process.

If the correlation function, and its second derivative, of a stationary random process $\xi(t)$ are continuous when $\tau = 0$, then in accordance with Section 8 this process will be continuous and differentiable.

Let us make use of formula (5.61) in order to find the first distribution function of the derivative of a normal random process. The starting point is the expression for the two-dimensional distribution function of process $\xi(t)$

$$w_2(x_1, x_2, \tau) = \frac{1}{2\pi\sigma^2\sqrt{1-R^2}} e^{-\frac{(x_1-a)^2 - 2R(x_1-a)(x_2-a) + (x_2-a)^2}{2\sigma^2(1-R^2)}}, \qquad (5.96)$$

where $R = R(\tau)$ is the correlation coefficient.

The distribution function $W(y, \tau)$ of the difference $\xi(t+\tau) - \xi(t)$ is equal to

$$W_1(y,\tau) = \int_{-\infty}^{\infty} w_2(u, y+u)\, du = \frac{e^{-\frac{y^2}{2\sigma^2(1-R^2)}}}{2\pi\sigma^2\sqrt{1-R^2}} \int_{-\infty}^{\infty} e^{-\frac{(u-a)^2 + y(u-a)}{\sigma^2(1+R)}} du =$$

$$= \frac{1}{2\pi\sigma^2\sqrt{1-R^2}} e^{-\frac{y^2}{2\sigma^2(1-R^2)}} e^{\frac{y^2}{4\sigma^2(1+R)}} \int_{-\infty}^{\infty} e^{-\frac{\left(u-a+\frac{y}{2}\right)^2}{\sigma^2(1+R)}} du,$$

and, noting that the integral is equal to $\sqrt{2\pi}\ \sigma\sqrt{\frac{1+R}{2}}$, we obtain

$$W_1(y,\tau) = \frac{1}{2\sqrt{\pi}\sigma\sqrt{1-R}} e^{-\frac{y^2}{4\sigma^2(1-R)}}. \qquad (5.97)$$

It is now not difficult to find the distribution function of

$$\frac{\xi(t+\tau) - \xi(t)}{\tau},$$

which is equal to

$$\tau W_1(y\tau,\tau)=\frac{\tau}{2\sqrt{\pi}\,\sigma\sqrt{1-R}}e^{-\frac{\tau^2y^2}{4\sigma^2(1-R)}}. \tag{5.98}$$

The desired distribution function of the derivative $W_1^{(\prime)}(y)$ is the limit of (5.98) when $\tau\to 0$. Since when $\tau\to 0$ $R(\tau)\to 1$, an indeterminacy results both in the numerator and the denominator. For resolution of this indeterminacy, we break down function $R(\tau)$ into a Maclaurin series

$$R(\tau)=R(0)+\tau R'(0)+\frac{\tau^2}{2}R''(0)+O(\tau^3).$$

But $R(0) = 1$, and $R'(0) = 0$ and $R''(0)<0$, since, when $\tau = 0$, the function $R(\tau)$ is at its maximum. Therefore

$$1-R(\tau)=-\frac{\tau^2}{2}R''(0)+O(\tau^3). \tag{5.99}$$

Substituting (5.99) into (5.98) and taking the limit, we obtain

$$W_1^{(1)}(\tau)=\lim_{\tau\to 0}\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\frac{1}{\sqrt{-R''(0)+O(\tau)}}e^{-\frac{y^2}{2\sigma^2[-R''(0)+O(\tau)]}},$$

wherefrom

$$W_1^{(1)}(\tau)=\frac{1}{\sqrt{2\pi\sigma^2\omega_1^2}}e^{-\frac{y^2}{2\sigma^2\omega_1^2}}, \tag{5.100}$$

where it is specified that

$$\omega_1^2=-R''(0),\ R''(0)<0. \tag{5.101}$$

Thus the one-dimensional distribution function of the derivative of a stationary normal process is normal, with a zero mean value and a dispersion of $\sigma^2\omega_1^2=\sigma^2|R''(0)|$.

It is possible to show that the two-dimensional distribution function of the derivative of a stationary normal process is also normal, and that its correlation function is $B_1(\tau) = -B''(\tau)$.

Thus the random process obtained by differentiating a stationary, normal process with a correlation function of $B(\tau)$ is stationary and normal, with a correlation function of $-B''(\tau)$.

Let us also define the joint distribution function of a stationary normal random process and its derivative in coincident instants of time. Employing (5.67') and (5.100), we find that

$$W_2(x, x') = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\sigma^2\omega_1^2}} e^{-\frac{x'^2}{2\sigma^2\omega_1^2}}. \tag{5.102}$$

The joint distribution of $\xi(t)$, $\xi'(t)$, $\xi(t+\tau)$, $\xi'(t+\tau)$ is also normal, and the determinant corresponding to this distribution (5.89) has the form [cf. (5.64) and (5.66)] of

$$D = \begin{vmatrix} 1 & 0 & R(\tau) & R'(\tau) \\ 0 & 1 & -R'(\tau) & -R''(\tau) \\ R(\tau) & -R'(\tau) & 1 & 0 \\ R'(\tau) & -R''(\tau) & 0 & 1 \end{vmatrix}. \tag{5.103}$$

13. Average Number of Intersections of Given Level in Unit of Time.

In solving a number of practical problems it is necessary to know the distribution function for the overshoot duration of a random process $\xi(t)$, where by overshoot duration is meant the segment of time, during which $\xi(t)$ exceeds an assigned level $x = x_0$. Sometimes instead of the distribution of overshoot duration one speaks of the null distribution of a random function, i.e., of the intersection of this function with an assigned horizontal straight line. In the general case there can be examined the problem of determining the distribution of the points of intersection of function $\xi(t)$ with an assigned function $f(t)$. A full-scale solution of the problem under consideration is very complex, but a number of useful approximations in special cases has been obtained by several authors [10 - 13].

Let us restrict ourselves to the simplest statistical characteristic of the overshoots of a random process — the average number of the intersection by this process of the assigned level $x = x_0$ in a unit of time.

The probability of intersecting level $x = x_0$ from below (i.e., with a positive derivative) with a sufficiently small $\Delta t > \frac{\Delta x}{\frac{d\xi}{dt}}$ coincides with the probability of the inequalities

$$x_0 - \Delta x < \xi(t) < x_0, \frac{d\xi(t)}{dt} > 0.$$

Let $w_2(x, y, t)$ be a two-dimensional distribution function of $\xi(t)$ and $\frac{d\xi(t)}{dt}$ at the coincident time instant $\underline{t}$.

Then

$$P\left\{x_0 - \Delta x < \xi(t) < x_0, \frac{d\xi(t)}{dt} > 0\right\} = \int_0^\infty \int_{x_0-\Delta x}^{x_0} w_2(x, y, t)\, dx dy.$$

With a sufficiently small Δt, the inside integral may be replaced by the expression $w_2(x_0, y, t)\Delta x = y w_2(x_0, y, t)\Delta t$ and

$$P\left\{x_0 - \Delta x < \xi(t) < x_0, \frac{d\xi(t)}{dt} > 0\right\} = \Delta t \int_0^\infty y w_2(x_0, y, t)\, dy. \tag{5.104}$$

The average number of the intersections of level $x = x_0$ during the time T is equal to

$$N_1(x_0, T) = \int_t^{t+T} P\left\{x_0 - \Delta x < \xi(t) < x_0, \frac{d\xi(t)}{dt} > 0\right\} = \int_t^{t+T}\int_0^\infty y w_2(x_0, y, t)\, dy dt, \tag{5.105}$$

and the average number of the intersections of this level in a unit of time (with a positive derivative) is equal to

$$n_1(x_0, T) = \frac{N_1(x_0, T)}{T} = \frac{1}{T}\int_t^{t+T}\int_0^\infty y w_2(x_0, y, t)\, dy dt. \tag{5.106}$$

For a stationary random process $\xi(t)$, a joint distribution of $\xi(t)$ and its derivative at a coincident time instant, t, does not depend upon t, and the mean number of intersections of the level $x = x_0$ per unit of time equals

$$n_1(x_0) = \int_0^\infty y w_2(x_0, y)\, dy. \tag{5.107}$$

With the employment of (5.67'), it is possible to represent $n_1(x_0)$ in the form

of

$$n_1(x_0) = w_1(x_0) \int_0^{\infty} y w_1^{(1)}(y)\, dy = m_1^{(+)} w_1(x_0), \tag{5.108}$$

where $m_1^{(+)}$ is an average of the positive values of derivative $\xi'(t)$.

Thus the curve for the dependence of the average number of overshoots on the level of x_0 coincides with the one-dimensional distribution curve of $\xi(t)$ with a precision given by a constant multiplier. So, for instance, the average number of overshoots of a normal random process is on the basis of (5.100) equal to

$$n_1(x_0) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x_0^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma\omega_1}} \int_0^{\infty} y e^{-\frac{y^2}{2\sigma^2\omega_1^2}} dy = \frac{\omega_1}{2\pi} e^{-\frac{x_0^2}{2\sigma^2}}. \tag{5.109}$$

In a completely analogous manner is obtained the formula for the average number of intersections, in a unit of time, of level $x = x_0$ from above (i.e., with a negative derivative)

$$n_2(x_0) = -\int_{-\infty}^{0} y w_2(x_0, y)\, dy = \int_{-\infty}^{0} |y|\, w_2(x_0, y)\, dy, \tag{5.110}$$

or

$$n_2(x_0) = \left| m_1^{(-)} \right| w_1(x_0). \tag{5.110'}$$

where $m_1^{(-)}$ is an average of the negative values of derivative $\xi'(t)$.

The overall average number of intersections of $\xi(t)$ with the straight line $x = x_0$ (the average number of nulls of a random function) is equal to the sum

$$n(x_0) = n_1(x_0) + n_2(x_0) = \int_{-\infty}^{\infty} |y|\, w_2(x_0, y)\, dy. \tag{5.111}$$

Literature

1. A. M. Yaglom, Vvedeniye v teoriyu statsionarnykh sluchaynykh funktsiy (Introduction to the theory of stationary random functions). "Uspekhi Matematicheskikh Nauk" ("Progress in the Mathematical Sciences"), 1952, V. VII, No. 5.

2. A. Ya. Khinchin. Teoriya korrelyatsii statisionarnykh sluchaynykh protsessov

(theory of the correlation of stationary random processes). "Uspekhi Matematicheskikh nauk", 1938, No. 5.

3. W. Meyer-Eppler, Korrelation und Autokorrelation in der Nachrichtentechnik. Archiv der Elektrischen Uebertragung, 7, No. 10, 1953; 7, No. 11, 1953.

4. Y. E. Lee, J. B. Wiesner, Correlation Functions and Communications Applications. Electrical Engineering, 73, No. 11, Nov. 1954.

5. A. Blanc-Lapierre, R. Fortet, Theorie des fonctions aleatoires. Paris, 1953.

6. V. S. Pugachev, Obshchaya teoriya korrelyatsii slychaynykh funktsiy (General theory of the correlation of random functions). Izv. AN SSSR, Math. Series, V. 17 No 5, 1953.

7. W. R. Bennett. Sources and Properties of Electrical Noise. Electrical Engineering, 73, No. 11, Nov. 1954.

8. W. B. Davenport, An Experimental Study of Speech-Wave Probability Distributions. Journ. Acoust. Sco. Amer. 24, 1952.

9. V.S. Pugachev. Elementy teorii sluchaynykh funktsiy, (Elements of the theory of random functions), VVIA (Military Air Engineering Academy) im Zhukovskogo, 1954.

10. D. Middleton, Spurious Signals Caused by Noise in Triggered Circuits. Journ. Appl. Phys. 19, Sept, 1948.

11. V. I. Bunimovich. Vybrosy napryazheniya fluktnatsionnykh shumov (Voltage overshoots of fluctuation noises). ZhTF, 1951, V. 21.

12. P. I. Kuznetsov, R. L. Stratonovich, B. I. Tikhonov. O dlitel'nosti vybrosov sluchaynoy funktsii (Overshoot duration of the random function). ZhTF, 1954, V. 24, No. 1.

13. V. I. Tikhonov. Raspredeleniye vybrosov normalinykh fluktuatsiy po dlitel'nosti (Overshoot distribution of normal fluctuations by length), "Radiotekhnika i elektronika" ("Radio Engineering and Electronics"), 1956, No. 1.

14. J. A. McFadden. The Axis-Crossing Intervals of Random Functions. Trans. IRE I T-2, No. 3, 1956.

15. F. H. Lange. Anwendung der Korrelationsanalyse in der Nachrichtentechnik. 1-VI.

Nachrichtentechni, No. 10, 1955, No. 1, No. 4, No. 7, No. 9, 1956. No. 1, 1957.

16. Fano. Short-Time-Autocorrelation Functions and Power Spectra. Journ. Acoust. Soc. Amer. No. 9, 1950.

17. A. A. Kharkevich. O priyeme slabykh signalov (The reception of weak signals). "Elektrosvyaz'" ("Electrical Communications"), 1956, No. 5.

18. D. L. Dub. Veroyatnostnyye Protsessy. Translation from English (cf. John L. Doob. "Stochastic Processes", N. Y., 1953) edited by A. M. Yaglom, Moskva, For. Lit. Pub. Hse. 1956.

19. I. M. Gel'fand. Obobshchennyye sluchaynyye protsessey (Generalized random processes). D. A. N. 100, No. 5, 1955.

20. M. S. Bartlett. An Introduction to Stochastic Processes, with Special Reference to Methods and Applications. Cambridge, 1955.

Chapter VI

TRANSFORMATIONS OF RANDOM PROCESSES

1. Definition and Terminology

A considerable number of radio-engineering problems consist in analyzing the results of the action of certain processes on devices of a greater or lesser degree of complexity. Setting aside a detailed characterization both of active processes and of radio-engineering systems, we shall only pause briefly on the basic definitions and terminology to the extent required for the forthcoming exposition.

An active process is the sum of a desired signal and of the noise distorting it; each of the terms of this sum may be either a given function of time, or a purely random process. In modern general communications theory, for instance, transmitted signals are regarded not as given functions of time, but as an aggregate of possible functions of time which have definite statistical properties.

The signal-distorting noise may be an internal noise of radio transmitting and receiving devices, or external noise (man-made, or atmospheric and industrial).

It is customary to divide elements of radio devices into two fundamental groups: nonlinear inertialess and linear (inertial or dynamic). The first group includes such elements as modulators, detectors, limiters, and mixers, the second — amplifiers and filters*.

The characteristic of a nonlinear, inertialess element is given in the form of the single-valued, non-linear transformation

$$y(t) = f\left[x(t)\right],$$

in which x refers to the input, and y to the output, the value of $y(t)$ at a given moment of time t being determined by the value of $x(t)$ only at the same moment of time t.

* It is known that such processes as modulation, detection and transformation may take place by means of linear systems with variable parameters. Within the framework of the present book, only linear systems with constant parameters will be examined.

A linear, inertial network is characterized by the transfer function $k(i\omega)$, the modulus and argument of which are respectively the frequency $c(\omega)$ and phase $\varphi(\omega)$ characteristics of this network.

In place of the frequency and phase characteristics, often the so-called transient pulse function $h(t)$ of a linear network is employed, linked with the transfer function by the Fourier transformation

$$h(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} k(i\omega)\, e^{i\omega t}\, d\omega. \qquad (6.1)$$

Employing the function $h(t)$, it is possible to express the process $y(t)$ at the output of a linear network in terms of the process at the input $x(t)$, by means of a relationship known as Duhamel's integral

$$y(t)=\int_{0}^{\infty} x(t-\tau)\, h(\tau)\, d\tau, \qquad (6.2)$$

from which it can be seen that the value of $y(t)$ at a given moment of time t depends on the values of x preceding the moment t. If, besides that, $x(t) \equiv 0$ when $t < 0$, then

$$y(t)=\int_{0}^{t} x(t-\tau)\, h(\tau)\, d\tau. \qquad (6.2')$$

The lower limit of integration in (6.2) is taken as equal to zero, and not to $-\infty$, since from the condition of the physical feasibility of the linear network, $h(t) \equiv 0$ when $t < 0$.

Strictly speaking, a nonlinear device (for instance, a detector) contains not only an element defined by a nonlinear functional relationship, but also a linear inertial element (the electrical circuit as detector load). On the other hand, the volt-ampere characteristics of linear networks preserve their linearity only within specific boundaries. Thus any element of radio apparatus should as a matter of principle be regarded as nonlinear and inertial. However, the existing mathematical apparatus is not in a condition to overcome those difficulties which arise in the

solution of problems with such general assumptions. Therefore the indicated division of radio apparatus into linear and nonlinear remains expedient for the time being. The error from such an idealization may be assessed for each concrete problem. Thus, for instance, if the plate load of a tube consists predominantly of resistance, the inertia of its characteristic may be neglected. A detector with a load in the form of a selective circuit may with a certain degree of approximation be regarded as a series connection of a nonlinear non-inertial network (rectifier) and a linear inertial network (load).

The total useful effect produced by a nonlinear device is obtained as a result of two transformations, of which one is nonlinear and non-inertial, while the other is linear and inertial.

In the following sections are examined the transformations of random processes in their passage through linear and nonlinear networks.

In this connection the problems examined can be of two types:

1) Determination of the correlation function and power spectrum of a process at the output of a network (minimum problem).

2) Determination of the multidimensional distribution functions of a process at the output of a network (maximum problem).

It is clear that from a solution of the second of the indicated problems a solution of the first can be obtained.

Actual radio apparatus consists of a whole series of linear and nonlinear elements. For this reason a complete solution of the problem will require several consecutive transformations of the random process.

A typical unit of radio apparatus is a link consisting of three elements: two linear networks, between which is located a nonlinear system (Fig. 43).

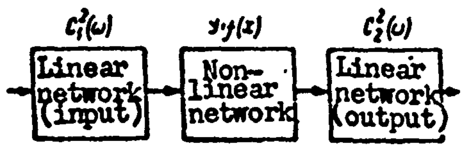


Fig. 43. Standard radio-apparatus link.

The transformations in such a standard link are characteristic of many stages of radio transmission and reception (frequency transformations, modulation, detection, limiting, etc.).

In this chapter, the general methods of solving the problems formulated above will be examined separately for linear and nonlinear networks. The application of these methods is illustrated in the following chapters. The question of the transformation of distribution functions in the standard link, if its input is acted upon by a normal random process, is examined separately in Chapter IX.

2. Power Spectrum and Correlation Function of Random Process at the Output of a Linear Network

In accordance with (6.2), a linear network with a transient pulse function, $h(\tau)$, transforms the random process $\xi_1(t)$, fed to its input, into another random process $\xi_2(t)$ which is an integral of the first

$$\xi_2(t)=\int_0^\infty \xi_1(t-\tau)\,h(\tau)\,d\tau. \qquad (6.3')$$

Let $F_1(\omega)$ be the power spectrum of the stationary random process $\xi_1(t)$ at the input of a linear network whose frequency characteristic is equal to $C(\omega)$. Then at the output of this network the power spectrum of process $\xi_2(t)$ has the form of*

$$F_2(\omega)=F_1(\omega)\ C^2(\omega). \qquad (6.4)$$

Formula (6.4) is the formula for the transformation of the power spectrum of a stationary random process in its passage through a linear network with a frequency characteristic of $C(\omega)$.

If the process at the input of the linear network consists of the sum of a stationary random process and a determined process, then, bearing in mind the principle of superposition as applicable to linear networks, it is possible to effect a

* A rigorous proof of formula (6.4) is obtainable from an examination of integral (6.3).

spectrum transformation according to (6.4) separately for each item, and then to sum up the results obtained.

As was to be expected, the phase characteristics of linear networks are in no way reflected in the transformation formula for the power spectrum of a process.

From (6.4) it follows that the correlation function $B_2(\tau)$ of a process at the output of a linear network is equal to

$$B_2(\tau)=\frac{1}{2\pi}\int_0^{\infty}F_1(\omega)C^2(\omega)\cos\omega\tau\,d\omega. \tag{6.5}$$

It is not difficult to obtain a different expression for $B_2(\tau)$, employing not the input spectrum $F_1(\omega)$, but the correlation function $B_1(\tau)$ of the process at the input.

Let us designate by $H(t)$ the Fourier transformation of $C^2(\omega)$, i.e.,

$$H(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}C^2(\omega)\,e^{i\omega t}\,d\omega.$$

Since

$$C^2(\omega)=k(i\omega)\,k(-i\omega),$$

therefore, with (6.1) in mind, we find on the basis of a theorem known in Fourier transformation theory as the convolution theorem, that

$$H(t)=\int_{-\infty}^{\infty}h(\tau)\,h(\tau+t)\,d\tau. \tag{6.6}$$

The inverse Fourier transformation of $F_1(\omega)$ is $B_1(\tau)$; therefore on the basis of the convolution theorem we obtain from (6.4) the expression for the correlation function of a random process at the output of a linear network

$$B_2(\tau)=\int_{-\infty}^{\infty}B_1(u)\,H(u+\tau)\,du. \tag{6.7}$$

Substituting for $H(t)$ its expression from (6.6), we obtain the link between the input and output correlation functions and the transient pulse function of the network

$$B_2(\tau)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}B_1(\tau+v-u)\,h(v)\,h(u)\,dv\,du. \tag{6.8}$$

From (6.5) and (6.8) there follows an equality useful in some computations

$$B_2(0)=\frac{1}{2\pi}\int_{0}^{\infty}F_1(\omega)\,C^2(\omega)\,d\omega= \\ =\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}B_1(v-u)\,h(v)\,h(u)\,dv\,du. \tag{6.8'}$$

Employing (6.7), the formula for the spectrum $F_2(\omega)$ of a process at the output of a network may be expressed in terms of the correlation function $B_1(\tau)$ of the process at the input

$$F_2(\omega)=4\int_{0}^{\infty}\int_{-\infty}^{\infty}B_1(u)\,H(u+\tau)\cos\omega\tau\,du\,d\tau. \tag{6.9}$$

Thus the problem of the transformation of the power spectrum of a random process and its correlation function in passage through linear networks is fully solved by formulas (6.4), (6.5), (6.8) and (6.9), if the frequency characteristic (or pulse function) of the network and the spectrum or correlation function of the process at the input are given.

In formula (6.3) it is tacitly assumed that the cut-in instant of $\xi_1(t)$ at the input of a linear network is equal to $-\infty$. If in such a case the random process $\xi_1(t)$ is stationary, random process $\xi_2(t)$ at the output of the linear network will also be stationary.

In cases when transients in linear networks are being studied, it is necessary to take into account the cut-in instant, i.e., to assume in (6.3) when $t<0$, that $\xi_1(t)\equiv 0$. The upper limit in integral (6.3) becomes variable and

$$\xi_2(t)=\int_{0}^{t}\xi_1(t-\tau)\,h(\tau)\,d\tau. \tag{6.10}$$

Thus the transient $\xi_2(t)$ in a linear network is a nonstationary random process, even if the random process $\xi_1(t)$ applied to its input is stationary. The average value

of the process at the output of the linear network in this case is equal to

$$a_2(t)=m_1\{\xi_2(t)\}=m_1\left\{\int_0^t \xi_1(t-\tau)h(\tau)d\tau\right\}=$$
$$=\int_0^t h(\tau)m_1\{\xi_1(t-\tau)\}d\tau,$$

and if $\xi_1(t)$ is a stationary random process, then

$$a_2(t)=a_1\int_0^t h(\tau)d\tau=a_1A(t)^*. \qquad (6.10')$$

The second distribution moment of the process in this case also is dependent on time, and in accordance with (6.8') is equal to

$$m_2\{\xi_2(t)\}=B_2(0,t)=\int_0^t\int_0^u B_1(v-u)h(v)h(u)dvdu, \qquad (6.10'')$$

where $B_1(\tau)$ is the correlation function of the stationary random process $\xi_1(t)$.

A special case of process $\xi_2(t)$ when $h(\tau)\equiv 1$, determined on the basis of (6.10), is the indefinite integral of the random process $\xi_1(t)$

$$\xi_2(t)=\int_0^t \xi_1(s)ds.$$

Since to the function $h(\tau)=1$ there corresponds $C(\omega)=\frac{1}{\omega}$, therefore the power spectrum for the integral of $\xi_1(t)$ is obtained by dividing its power spectrum by ω^2, i.e.,

$$F_2(\omega)=\frac{1}{\omega^2}F_1(\omega). \qquad (6.11)$$

Let us note that, employing the filtration properties of the derivatives of the delta-function (cf. Appendix IV), it is possible to extend Duhamel's integral (6.3) to differentiation processes of random functions. Thus, for instance, the first derivative of process $x(t)$ may be regarded as a result of the passage of this process through a linear network with a transient pulse function equal to $\delta^{(1)}(t)$. It then follows from (6.6), that for such a linear network $H(t)=\delta^{(2)}(t)$, and the square of the frequency characteristic $C^2(\omega)=\omega^2$.

---

* The function $A(t)=\int_0^t h(\tau)\,d\tau$ is called the transient conductance of a linear network or the reaction of a linear network to a unit step.

Employing the general formula (6.4) for the transformation of the power spectrum of a random process in linear networks, we find the spectrum of the derivative $F_2(\omega) = \omega^2 F_1(\omega)$ that corresponds to formula (5.65) which was obtained on the basis of other considerations.

3. Passage of White Noise through a Linear Network.

Let us examine the case, of practical importance, when on the input of a linear network there acts a random process with a power spectrum uniform for all frequencies (white noise) (cf. 5.83)*.

$$F_1(\omega) = F_0 = \text{const.}$$

Then in accordance with (6.4) the power spectrum $F_2(\omega)$ of a random process at the output of this network is equal to

$$F_2(\omega) = F_0 C^2(\omega). \qquad (6.12)$$

In this manner, the power spectrum of white noise at the output of a linear network coincides in shape with the square of the frequency characteristic of the network.

From (6.12) we obtain directly the correlation function of a random process at the output of a linear network

$$B(\tau) = \frac{F_0}{2\pi} \int_0^\infty C^2(\omega) \cos \omega\tau \, d\omega. \qquad (6.13)$$

Bearing in mind (6.5) and (6.6), we find

$$B(\tau) = \frac{1}{2} F_0 H(\tau) = \frac{1}{2} F_0 \int_{-\infty}^{\infty} h(u) h(u+\tau) \, du. \qquad (6.13')$$

Thus the correlation function of white noise at the output of a linear network coincides, with an accuracy of to the last constant factor, with the convolution of the transient pulse function of the network.

* An example of white noise is the fluctuation noises of radio receivers. The intensity F_0 of these noises per unit of frequency band is equal to $n_{ш} kT$, where k is Boltzmann's constant, T is the absolute temperature ($kT \approx 4 \cdot 10^{-21}$ watt/cps) and $n_{ш}$ is the noise coefficient of the receiver.

For a narrow-band linear network with a symmetrical frequency characteristic relative to a high frequency ω_0, the correlation function at the output may, in accordance with (5.82), be represented in the form of

$$B(\tau)=a(\tau)\cos\omega_0\tau=\left[\frac{F_0}{\pi}\int\limits_0^\infty C_1^2(\omega)\cos\omega\tau d\omega\right]\cos\omega_0\tau, \tag{6.14}$$

where
$$C_1(\omega)=C(\omega_0-\omega).$$

It can be seen from (6.14), that the correlation function of white noise which has passed through a narrow-band linear network with a resonance frequency of ω_0, is equal to the product of $\cos\omega_0\tau$ and the correlation function $a(\tau)$ of white noise which has passed through a network with a frequency characteristic of $C_1(\omega)$, obtained by displacing $C(\omega)$ into the low-frequency range by the amount ω_0.

With the passage of white noise through linear networks is connected the so-called power determination of the width of the pass band of their frequency characteristics. According to this determination, the width of the pass band is computed from the full power of the white noise passing through a linear network, the actual frequency characteristic being replaced by an idealized rectangular characteristic, equivalent to the power of the noise (Fig. 44). The height of the rectangle is so selected, that the constant density of the power spectrum within the limits of its width equals the maximum density $F(\omega_0)$ of the actual power spectrum at the output of the linear network. In accordance with the assumed determination the width of the pass band, Δf, of the frequency characteristic $C(\omega)$ is computed by the formula

$$\Delta f=\frac{\frac{1}{2\pi}\int\limits_0^\infty F(\omega)\,d\omega}{F(\omega_0)}, \tag{6.15}$$

or, taking into account (6.12),

$$\Delta f=\frac{\int\limits_0^\infty C^2(\omega)\,d\omega}{2\pi C^2(\omega_0)}. \tag{6.16}$$

From (6.15) and (5.46) there follows also

$$\Delta f = \frac{B(0)}{F_0 C^2(\omega_0)}. \quad (6.16')$$

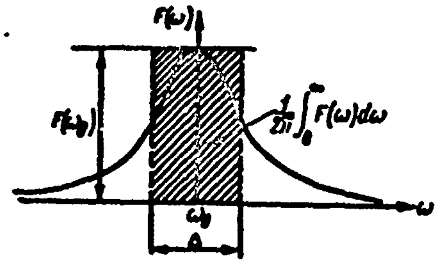


Fig. 44. Power determination of width of pass band of a frequency characteristic.

In practice it is frequently convenient to measure the bandwidth of a linear network between the points at which the amplification in terms of power is equal to one half of its maximum value, or the amplification in terms of voltage constitutes 0.7 of the maximum. For theoretical investigations it is more convenient to employ the power determination of the bandwidth in the form of (6.15). However, both methods give close results in the most important practical cases.

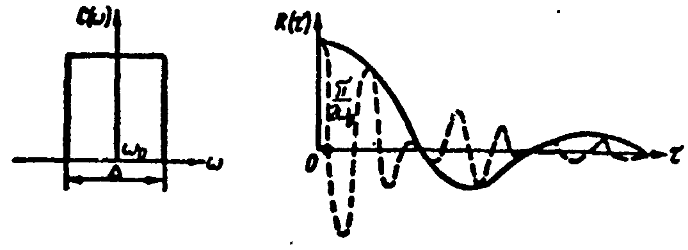
The correlation time of white noise which has passed through a linear network may be determined in the following manner [compare (5.32)] :

$$\tau_0 = \frac{\int_0^\infty a(\tau)\,d\tau}{B(0)} = \frac{F_0 C_1^2(0)}{2B(0)}, \quad (6.17)$$

or, taking into account (6.16'), $\tau_0 = \frac{1}{2\Delta f}$. In this manner, the correlation time has an order of magnitude inverse to the bandwidth of the linear network.

Let us examine several examples of the passage of white noise through linear networks with frequency characteristics of a definite form.

a) The passage of white noise through an ideal linear network (filter or amplifier) with a pass band equal to $\Delta = 2\pi\Delta f$ (Fig. 45, a).



a) b)

Fig. 45. a) Frequency characteristic of ideal linear network.

b) Correlation function of white noise at output network.

The equation for the frequency characteristic of an ideal linear network has the form of

$$C(\omega)=\begin{cases} C_0, & |\omega-\omega_0|<\frac{\Delta}{2}, \\ 0, & |\omega-\omega_0|>\frac{\Delta}{2}, \end{cases}$$

and, consequently, the power spectrum of noise at the output of this filter is equal to

$$F(\omega)=\begin{cases} F_0C_0^2, & |\omega-\omega_0|<\frac{\Delta}{2} \\ 0, & |\omega-\omega_0|>\frac{\Delta}{2}. \end{cases}$$

When $\Delta \ll \omega_0$, the noise at the output of an ideal linear network represents an example of a process whose power spectrum lies in a narrow frequency band about a high frequency ω_0, and is symmetrical with respect to the latter. Then in accordance with (6.14), the correlation function of the noise at the output of an ideal linear network is equal to

$$B(\tau)=\frac{F_0}{\pi}\cos\omega_0\tau\int\limits_0^{\frac{\Delta}{2}} C_0^2\cos\omega\tau d\omega=\frac{F_0C_0^2}{\pi}\frac{\sin\frac{\tau\Delta}{2}}{\tau}\cos\omega_0\tau.$$

The full power of the noise is

$$B(0)=\frac{F_0C_0^2}{2\pi}\Delta,$$

and, consequently, the expression for the correlation coefficient of the noise at the output of an ideal filter (amplifier) has the form of

$$R(\tau) = \frac{\sin\frac{\tau\Delta}{2}}{\frac{\tau\Delta}{2}} \cos \omega_0 \tau. \tag{6.18}$$

In Figure 45b, the curve of correlation function (6.18) is indicated by the dotted line. The curve is plotted only for the positive argument of τ; in virtue of the evenness of the correlation coefficient the ordinate axis is its axis of symmetry. The solid curve corresponds to the limiting case of $\omega_0 = 0$, i.e., to the correlation function of noise at the output of a low-frequency filter (amplifier) with a bandwidth of $\frac{\Delta}{2}$.

If the correlation time is determined according to inequality (5.32'), it follows from (6.18) that the instantaneous amplitudes of the noise, separated by a time interval of

$$\tau > \frac{40}{\Delta} = \frac{6.4}{\Delta f},$$

may be considered as being practically independent (cf. p. 180).

If, however, the correlation time is computed on the basis of (6.17), then

$$\tau_0 = \int_0^\infty \frac{\sin\frac{\tau\Delta}{2}}{\frac{\tau\Delta}{2}} d\tau = \frac{\pi}{\Delta} = \frac{1}{2\Delta f}.$$

Let us also find the dispersion σ_1^2 of the derivative of white noise which has passed through an ideal network with a resonance frequency of ω_0 and a bandwidth of Δ. In accordance with (5.64'), $\sigma_1^2 = -B''(0)$.

For the computation of $B''(0)$ it is simplest of all to expand $B(\tau)$ into a series in terms of τ, restricting one's self to terms of no higher than the second power

$$B(\tau) = B(0) \frac{\sin\frac{\tau\Delta}{2}}{\frac{\tau\Delta}{2}} \cos \omega_0 \tau = B(0)\left(1 - \frac{\tau^2\Delta^2}{4\cdot 6} + \ldots\right) \times$$

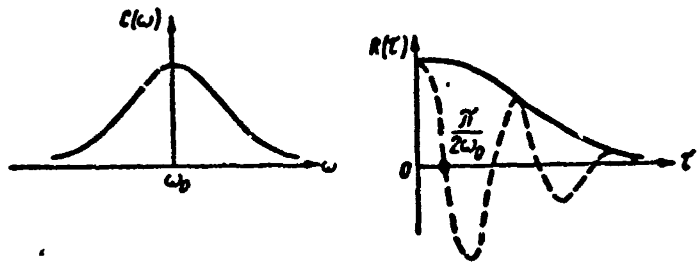
$$\times \left(1 - \frac{\omega_0^2\tau^2}{2} + \ldots\right) = B(0)\left(1 - \frac{\tau^2\Delta^2}{24} - \frac{\omega_0^2\tau^2}{2} + \ldots\right).$$

Then

$$\sigma_1^2 = -B''(0) = B(0)\left(\frac{\Delta^2}{12} + \omega_0^2\right) = \frac{F_0 C_0^2}{2\pi}\Delta\left(\frac{\Delta^2}{12} + \omega_0^2\right).$$

b) Passage of white noise through a linear network (multistage amplifier) with a frequency characteristic in the form of a gaussian curve (Fig. 46,a)

$$C(\omega) = C_0 e^{-\frac{(\omega-\omega_0)^2}{2\beta^2}}.$$



a) b)

Fig. 46. a) Frequency characteristic in the form of gaussian curve.

b) Correlation function of white noise at output of network.

The power spectrum of the noise at the output of such a network also has the form of a gaussian curve

$$F(\omega) = F_0 C_0^2 e^{-\frac{(\omega-\omega_0)^2}{\beta^2}}.$$

The correlation function of the noise is, in accordance with (6.14), equal to

$$B(\tau) = \frac{F_0 C_0^2}{\pi}\cos\omega_0\tau\int_0^\infty e^{-\frac{\omega^2}{\beta^2}}\cos\omega\tau\, d\omega =$$

$$= \frac{F_0 C_0^2}{2\sqrt{\pi}}\beta e^{-\frac{\beta^2\tau^2}{4}}\cos\omega_0\tau.$$

The full power of the noise is

$$B(0) = \frac{F_0 C_0^2}{2\sqrt{\pi}}\beta.$$

and, consequently, the expression for the correlation coefficient of the noise in the case at hand has the form of

$$R(\tau)=e^{-\frac{\beta^2\tau^2}{4}}\cos\omega_0\tau. \tag{6.19}$$

The curve of correlation function (6.19) is shown in Figure 46,b. Employing formula (6.16) we find, for the case under consideration, the power pass bandwidth $\Delta = \beta\sqrt{\pi}$.

In this manner, the parameter β in the equation of the frequency characteristic coincides, with an accuracy of constant factor $\sqrt{\pi}$, with the pass bandwidth.

The dispersion of the derivative of white noise, which has passed through a linear network with a gaussian frequency characteristic, is equal to

$$\sigma_1^2=-B''(0)=B(0)\left(\frac{\beta^2}{2}+\omega_0^2\right)=\frac{F_0C_0^2}{2\pi}\Delta\left(\frac{\Delta^2}{2\pi}+\omega_0^2\right).$$

Linear networks with frequency characteristics of the two types examined above are an idealization, and are physically unfeasible. It is possible to examine physically feasible linear networks, whose frequency characteristics approach those examined above. Some standard examples of such frequency characteristics are cited in [4] .

c) Passage of white noise through an oscillating circuit formed by the parallel connection of inductance L, resistance R, and capacitance C.

As is known, the frequency characteristic of the oscillating circuit under examination has (Fig. 47,a) the form of

$$C(\omega)=\frac{\omega_0^2}{\sqrt{(\omega^2-\omega_0^2)^2+4\alpha^2\omega^2}},$$

where

$$\alpha=\frac{1}{2RC},\quad \omega_0=\frac{1}{\sqrt{LC}}.$$

The power noise spectrum at the output of this oscillating circuit is defined by the function

$$F(\omega)=\frac{F_0\omega_0^4}{(\omega^2-\omega_0^2)^2+4\alpha^2\omega^2}.$$

If $\omega_0 \gg \alpha$, then with a shift of the frequency characteristic into the low-frequency range by the amount ω_0, its equation will be transformed into the form of

$$C_1(\omega) = \frac{\omega_0}{2\sqrt{\omega^2 + \alpha^2}}.$$

Then in accordance with (6.14) the correlation function of noise at the output of the oscillating circuit is equal to

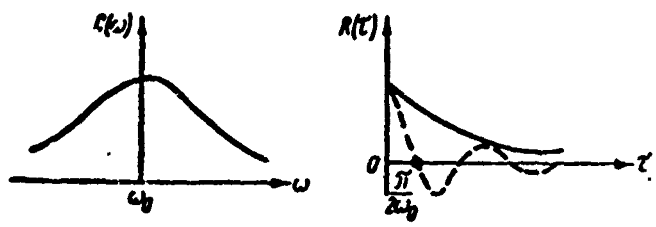
$$B(\tau) = \frac{F_0 \omega_0^2}{4\pi} \cos \omega_0 \tau \int_0^\infty \frac{\cos \omega\tau}{\omega^2 + \alpha^2}\, d\omega = \frac{F_0 \omega_0^2}{8\alpha} e^{-\alpha|\tau|} \cos \omega_0 \tau.$$

The full power of the noise is

$$B(0) = \frac{F_0 \omega_0^2}{8\alpha},$$

and, consequently, the expression for the correlation coefficient of noise has the form of*

$$R(\tau) = e^{-\alpha|\tau|} \cos \omega_0 \tau. \qquad (6.20)$$



a) b)

Fig. 47. a) Frequency characteristic of oscillating circuit.

b) Correlation function of white noise at output of circuit.

The curve of correlation function (6.20) is shown in Figure 47,b. Employing formula (6.16), we find the power pass bandwidth of the oscillating circuit frequency characteristic $\Delta = \pi\alpha$.

Let us note that the curves of the correlation function and power spectrum, of

* As can be seen from (6.20), the derivatives of correlation function (6.20) undergo a break when $\tau = 0$. This indicates that, for the random process observed at the output of an oscillating circuit when a white noise is injected at its input, a time derivative does not exist. For more details, see in [14].

white noise which has passed through an oscillating circuit, do not differ in type from the corresponding curves for a generalized telegraph signal (cf. Fig. 41).

With an arbitrary relationship between ω_0 and α, formula (6.14) ceases to be valid, and to find a correlation function it is necessary to compute (for example - with the aid of residues theory) the integral

$$\int_0^\infty \frac{\cos\omega\tau\, d\omega}{(\omega^2-\omega_0^2)^2+4\alpha^2\omega^2}.$$

As a result of the computation of the indicated integral, there are obtained the following formulas for the correlation coefficient of noise at the output of a circuit:

when $\omega_1^2 = \omega_0^2 - \alpha^2 > 0$ (oscillating circuit)

$$R(\tau)=e^{-\alpha|\tau|}\left(\cos\omega_1\tau+\frac{\alpha}{\omega_1}\sin\omega_1|\tau|\right); \qquad (6.21)$$

when $\omega_0 = \alpha$ (boundary of oscillating and aperiodic circuits)

$$R(\tau)=e^{-\alpha|\tau|}(1+\alpha|\tau|); \qquad (6.22)$$

when $\beta^2 = \alpha^2 - \omega_0^2 > 0$ (aperiodic circuit)

$$R(\tau)=\frac{1}{2}\left[\left(1+\frac{\alpha}{\beta}\right)e^{-(\alpha-\beta)|\tau|}+\left(1-\frac{\alpha}{\beta}\right)e^{-(\alpha+\beta)|\tau|}\right]. \qquad (6.23)$$

For an arbitrary linear network, the correlation function of noise at the output of this system will consist of a sum of terms of the type of (6.21) and (6.23), the total number of which is equal to half the total number of poles of the transfer function of the linear circuit under examination.

4. Optimum Linear Networks.

A problem of practical importance is the determination of such a linear network (filter), which in the best possible manner separates a signal from noise. In distinction from Sect. 5 Ch. V, we shall examine a more general problem, when the signal $\xi(t)$ and noise $\eta(t)$ distorting it, are independent random stationary functions of time, the correlation functions and power spectra of which are known and equal $B_c(\tau)$,

$F_c(\omega)$, and $B_n(\tau)$, $F_n(\omega)$ respectively.

The sum of $\xi(t) + \eta(t)$ is fed into the input of a linear network, and at its output there is received the random process $\zeta(t)$. The optimum separation of signal from error is given by a linear network which realizes the minimum mean-square deviation in time of the process at the output of the linear network from the signal $\xi(t)$, i.e., the magnitude

$$\varepsilon(t_0)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[\zeta(t)-\xi(t+t_0)]^2\,dt. \tag{6.24}$$

The possible time shift, $t_0>0$, is examined in problems dealing with the linear prediction of a signal.

The definition of the optimum criterion is, of course, conditional and depends on the methods of observing the output process. There can be problems where the optimum signal separation must satisfy a condition different from that of (6.24). The indicated criterion is usually adopted for automatic control systems.

Employing (6.3), it is possible to represent the expression for mean-square error in the form of*

$$\varepsilon(t_0)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left\{\xi(t+t_0)-\int_0^{\infty}[\xi(t-\tau)+\eta(t-\tau)]\,h(\tau)\,d\tau\right\}^2 dt. \tag{6.25}$$

Taking into account the independence of $\xi(t)$ and $\eta(t)$, and also bearing in mind that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\xi^2(t+t_0)\,dt=B_c(0),$$

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\xi(t+t_0)\,\xi(t-\tau)\,dt=B_c(t_0+\tau),$$

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\left\{\int_0^{\infty}[\xi(t-\tau)+\eta(t-\tau)]\,h(\tau)\,d\tau\right\}^2 dt=$$

* It is, consequently, assumed that the action of the process on the input of the filter begins in distant instants of time, and that the process at the output is also stationary.

$$=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\int_0^\infty\int_0^\infty[\xi(t-u)+\eta(t-u)]\times$$
$$\times[\xi(t-v)+\eta(t-v)]h(u)h(v)\,du\,dv\,dt=$$
$$=\int_0^\infty\int_0^\infty[B_c(u-v)+B_n(u-v)]h(u)h(v)\,du\,dv,$$

we find

$$\varepsilon(t_0)=B_c(0)-2\int_0^\infty B_c(t_0+\tau)h(\tau)\,d\tau+$$
$$+\int_0^\infty\int_0^\infty[B_c(u-v)+B_n(u-v)]h(u)h(v)\,du\,dv. \qquad (6.26)$$

Thus the mean-square error in the reproduction of a signal at the output of a linear network depends only on the mean power characteristics of the signal and noise i.e., on their correlation functions. The various signals $\xi_K(t)$, which have the same correlation functions, will be separated from the noise $\eta_K(t)$, likewise with uniform correlation functions, in the best manner by one and the same linear network.

We shall show that with $B_c(\tau)$ and $B_n(\tau)$ given, the best separation of signal from noise will be accomplished by such a linear network, the pulse function, $h^*(t)$, of which satisfies the integral equation

$$B_c(\tau)=\int_0^\infty[B_c(\tau-u)+B_n(\tau-u)]h^*(u)\,du. \qquad (6.27)$$

We shall substitute (6.27) into (6.26) and prove, that the error of reproduction is at a minimum only under the condition that $h(t)\equiv h^*(t)$. In actual fact, as a result of such a substitution we have

$$\varepsilon=B_c(0)-2\int_0^\infty\int_0^\infty[B_c(\tau-u)+B_n(\tau-u)]h^*(u)h(\tau)\,du\,d\tau+$$
$$+\int_0^\infty\int_0^\infty[B_c(u-v)+B_n(u-v)]h(u)h(v)\,du\,dv=$$
$$=B_c(0)-\int_0^\infty\int_0^\infty[B_c(\tau-u)+B_n(\tau-u)]h^*(u)h(\tau)\,du\,d\tau-$$
$$-\int_0^\infty\int_0^\infty[B_c(\tau-u)+B_n(\tau-u)]h(u)h^*(\tau)\,du\,d\tau+$$

$$+\int_0^\infty\int_0^\infty [B_c(u-v)+B_n(u-v)]\,h(u)\,h(v)\,du\,dv=$$

$$=B_c(0)-\int_0^\infty\int_0^\infty [B_c(\tau-u)+B_n(\tau-u)]\,h^*(u)\,h^*(\tau)\,du\,d\tau+$$

$$+\int_0^\infty\int_0^\infty [B_c(u-v)+B_n(u-v)]\,[h(u)-h^*(u)]\,[h(v)-h^*(v)]\,du\,dv.$$

Employing (6.8'), it is possible to rewrite the obtained expression in the form of

$$\epsilon=B_c(0)-\frac{1}{2\pi}\int_0^\infty [F_c(\omega)+F_n(\omega)]\,|k^*(i\omega)|^2\,d\omega+$$
$$+\frac{1}{2\pi}\int_0^\infty [F_c(\omega)+F_n(\omega)]\,|k(i\omega)-k^*(i\omega)|^2\,d\omega, \qquad (6.28)$$

where $k(i\omega)$ and $k^*(i\omega)$ are transfer functions of linear networks, corresponding to the pulse functions $h(t)$ and $h^*(t)$. Since only the third term in (6.28) contains $k(i\omega)$ and since it is essentially positive, therefore the minimum value of ϵ will correspond to such a linear network, the transfer function $k(i\omega)$ of which turns the third term in (6.28) to zero. As is not difficult to see, this will take place under the condition that $k(i\omega)\equiv k^*(i\omega)$, Q.E.D.

It follows from (6.28), that the minimum value of the mean-square error of signal reproduction at the output of an optimum linear network is equal to

$$\epsilon_{min}=B_c(0)-\frac{1}{2\pi}\int_0^\infty [F_c(\omega)+F_n(\omega)]\,|k^*(i\omega)|^2\,d\omega=$$
$$=\frac{1}{2\pi}\int_0^\infty \{F_c(\omega)-[F_c(\omega)+F_n(\omega)]\,|k^*(i\omega)|^2\}\,d\omega. \qquad (6.29)$$

The fundamental difficulty in solving the integral equation (6.27) is linked with the fact that function $h^*(t)$ must satisfy the condition of physical feasibility (cf. p. 217), as a result of which the lower limit of integration is equal to zero, and not to $-\infty$. If the condition of physical feasibility is discarded and the integration is extended from $-\infty$ to $+\infty$, then the integral equation is easily solved by means of a Fourier transformation of both parts of (6.27). As the result of such a transformation we obtain a solution in the form of

$$k^*(i\omega)=\frac{F_c(\omega)}{F_c(\omega)+F_n(\omega)}.$$

from which we find the expression for the frequency characteristic of the optimum linear network*

$$C^{*}(\omega)=\frac{F_c(\omega)}{F_c(\omega)+F_n(\omega)}. \tag{6.30}$$

The mean quadratic error in the reproduction of a signal at the output of a linear network is, in accordance with (6.29), equal in this case to

$$\varepsilon_{min}=\int_{-\infty}^{\infty}\frac{F_c(\omega)\cdot F_n(\omega)}{F_c(\omega)+F_n(\omega)}\,d\omega. \tag{6.31}$$

Formula (6.31) indicates that an error in the reproduction of a signal in the optimum linear network may be made equal to zero only in the case when the spectra of the signal and noise do not overlap, i.e., when $F_c(\omega)\cdot F_n(\omega)\equiv 0$. In the contrary case an error is inevitable.

In the works [2], [3] there is presented a complete solution of the problem examined above, with allowance for the physical feasibility of linear networks, including the case when the signal and noise are correlated.

In some cases the amount of the mean-square error ε_{min} can be reduced, if the pass band of the linear network varies with the statistical properties of the input random process [12].

5. Envelope and Phase of a Random Process.

As is known from the theory of the Fourier integral (cf., for instance, E. Titchmarsh, Vvedeniye v teoriyu integrala Fur'e. Goskekhizdat, 1948 [(i.e., Introduction to the Theory of the Fourier Integral (2nd edition, New York, 1948)]), by complying with certain conditions, the function S(t) may be presented in the form of $S(t) = a(t)\cos\varphi(t)$.

Functions a(t) and $\varphi(t)$ are determined on the basis of the formulas

$$a(t)=\sqrt{S^2(t)+\sigma^2(t)},\quad \varphi(t)=\operatorname{arctg}\frac{\sigma(t)}{S(t)}.$$

* The phase characteristic in such a solution does not depend on frequency, and constitutes a constant which is a multiple of 2π.

where $\sigma(t)$ is a so-called conjugate function of $S(t)$, which is uniquely defined on the basis of $S(t)$ by the Gilbert integral transformation

$$\sigma(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{S(\tau)}{\tau - t}\,d\tau.$$

Analogously, with several very general assumptions it is possible, on the basis of a given random stationary process $\xi(t)$, to form by means of a Gilbert transformation the conjugate of $\xi(t)$, the new stationary random process $\eta(t)$

$$\eta(t) = -\frac{1}{\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{\xi(\tau)}{\tau - t}\,d\tau. \tag{6.32}$$

Then the initial random process $\xi(t)$ may be represented in the form of

$$\xi(t) = E(t)\cos\Phi(t), \tag{6.33}$$

while the random processes $E(t)$ and $\phi(t)$, being expressed in terms of $\xi(t)$ and of its conjugate $\eta(t)$, in the formulas

$$E(t) = \sqrt{\xi^2(t) + \eta^2(t)}, \tag{6.34}$$

$$\Phi(t) = \operatorname{arc\,tg}\frac{\eta(t)}{\xi(t)}. \tag{6.35}$$

The random processes $E(t)$ and $\phi(t)$, defined in this manner, are called respectively the envelope and the phase of the random process $\xi(t)$.

Let us note that from (6.34) it follows that $E(t) \geqslant \xi(t)$, i.e., the random function $\xi(t)$ nowhere intersects the random function $E(t)$. In addition

$$E\frac{dE}{dt} = \xi\cdot\xi' + \eta\cdot\eta',$$

and therefore at points where $\xi(t) = E(t)$, (i.e., $\eta(t) = 0$), there holds true $\frac{dE}{dt} = \frac{d\xi}{dt}$. Thus the random function $\xi(t)$ does not intersect $E(t)$, and at points of contact has a common tangent. The indicated properties explain the sense of the accepted designation of random function $E(t)$ as the envelope of $\xi(t)$.

When the random process $\xi(t)$ is represented in the form of (6.33), it may be regarded as a harmonic vibration, modulated with respect to amplitude and phase by

the random functions E(t) and ϕ(t).

The possibility of such a representation imposes no essential restrictions on the power spectrum of the process. However, the representation in question assumes the greatest practical and graphic interest for narrow-band processes (cf. Section 10, Chapter V). With a small width of the power spectrum of process ξ(t), its envelope E(t) varies relatively slowly in time (as compared with cos φ(t)), i.e., the values of E(t) are strongly correlated and its correlation function varies slowly with respect to τ, and its power spectrum will be concentrated principally in the low-frequency range. If ω_0 is the central frequency with respect to which the narrow-band spectrum ξ(t) is situated, then the phase ϕ(t) is equal to

$$\Phi(t)=\omega_0 t-\varphi(t), \tag{6.36}$$

the random function φ(t) also turning out to vary slowly in time.

Substituting (6.36) into (6.33), we obtain the following representation of the narrow-band random process:

$$\xi(t)=E(t)\cos\varphi(t)\cos\omega_0 t+E(t)\sin\varphi(t)\sin\omega_0 t, \tag{6.37}$$

and designating

$$A(t)=E(t)\cos\varphi(t),\quad C(t)=E(t)\sin\varphi(t), \tag{6.38}$$

we obtain

$$\xi(t)=A(t)\cos\omega_0 t+C(t)\sin\omega_0 t. \tag{6.39}$$

Thus the narrow-band random process is characterized as a high-frequency oscillation with a carrier frequency of ω_0 and with slowly varying envelope and phase.

It is possible to show [13], that the correlation function and power spectrum of the random process η(t), conjugate of ξ(t), coincides with the correlation function, B(τ) and the power spectrum F(τ) of the random process ξ(t), and that the mutual correlation functions of these processes are obtained from the relationship

$$m_1\{\xi(t)\eta(t+\tau)\} = -m_1\{\eta(t)\xi(t+\tau)\} = \tag{6.40}$$
$$= \frac{1}{2\pi}\int_0^\infty F(\omega)\sin\omega\tau\,d\omega.$$

Since

$$\eta(t) = E(t)\sin\Phi(t) = A(t)\sin\omega_0 t - C(t)\cos\omega_0 t, \tag{6.41}$$

it follows from (6.39) and (6.41) that

$$A(t) = \xi(t)\cos\omega_0 t + \eta(t)\sin\omega_0 t, \tag{6.42}$$
$$C(t) = \xi(t)\sin\omega_0 t - \eta(t)\cos\omega_0 t. \tag{6.43}$$

Let us designate by $B_A(\tau)$, $B_C(\tau)$, $B_{AC}(\tau)$ and $B_{CA}(\tau)$ the correlation and mutual correlation function of the random processes A(t) and C(t). Then from (6.42) it follows that

$$B_A(\tau) = B_C(\tau) = m_1\{\xi(t)\xi(t+\tau)\}\cos\omega_0\tau +$$
$$+ m_1\{\xi(t)\eta(t+\tau)\}\sin\omega_0\tau,$$
$$B_{AC}(\tau) = -B_{CA}(\tau) = m_1\{\xi(t)\xi(t+\tau)\}\sin\omega_0\tau -$$
$$- m_1\{\xi(t)\eta(t+\tau)\}\cos\omega_0\tau,$$

or, expressing the correlation and mutual correlation functions of process ξ(t) and η(t) in terms of the power spectrum F(ω) of process ξ(t), we obtain

$$B_A(\tau) = B_C(\tau) = \frac{1}{2\pi}\int_0^\infty F(\omega)\cos(\omega-\omega_0)\tau\,d\omega, \tag{6.44}$$

$$B_{AC}(\tau) = -B_{CA}(\tau) = \frac{1}{2\pi}\int_0^\infty F(\omega)\sin(\omega-\omega_0)\tau\,d\omega. \tag{6.45}$$

If the power spectrum F(ω) of the random process ξ(t) has a narrow band and is symmetrical with respect to the central frequency ω_0, then from (6.44) and (6.45) it follows (cf. Section 10, Chapter V) that

$$B_A(\tau) = B_C(\tau) = \frac{1}{\pi}\int_0^\infty F^*(\omega)\cos\omega\tau\,d\omega = a(\tau), \tag{6.46}$$

$$B_{AC}(\tau) = -B_{CA}(\tau) = 0. \tag{6.47}$$

where a(τ) is a slowly changing function of argument τ.

From (6.39) and (6.46) it follows that the correlation function B(τ) of the narrow-band random process ξ(t) is equal to

$$B(\tau)=m_1\{\xi(t)\xi(t+\tau)\}=a(\tau)\cos\omega_0\tau, \tag{6.48}$$

which coincides with (5.82). Here, however, becomes clear the physical sense of the function $a(\tau)$ in (5.82), which is the correlation function for each of the slowly varying processes A(t) and C(t), tied to the narrow-band process $\xi(t)$ by the relationship (6.39).

As was seen from Section 3, white noise, which has passed through a narrow-band linear network with a resonance frequency of ω_0, constitutes a narrow-band random process. If, besides that, white noise is normal, then the process at the output of the narrow-band linear network may be represented as a sum of the type of (6.39), in which A(t) and C(t) will be incoherent normal and, consequently, independent random processes.

6. Correlation Function and Power Spectrum of Random Process at Output of Nonlinear Network.

In Section 3 it was indicated that in linear systems the spectrum and correlation function are uniquely determined by the frequency characteristic of the network and by the spectrum of the process at its input. In order to determine the power spectrum of a random process or of its correlation function, it is insufficient to know only the spectrum (or the correlation function) at its input, and it is necessary also to have the expression for the two-dimensional distribution function of the input random process.

Let the characteristic of a nonlinear and noninertial network have the form of

$$y=f(x) \tag{6.49}$$

and let there be known the second distribution function $\omega_2(x_1, x_2, \tau)$ of the stationary random process $\xi(t)$ at the input of the nonlinear network. Then, employing the rules cited in Chapter III for finding mean values of the functions of random variables, for the correlation function of a random process at the output of a nonlinear network we obtain the following expression:

$$B(\tau)=m_1\{f[\xi(t)]f[\xi(t+\tau)]\}=$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x_1)f(x_2)w_2(x_1,x_2,\tau)\,dx_1\,dx_2. \tag{6.50}$$

The mean value of a (the direct component) is, in accordance with (5.23), determined on the basis of the asymptotic value of $B(\tau)$, i.e.,

$$a^2=\lim_{\tau\to\infty}B(\tau)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x_1)f(x_2)[\lim_{\tau\to\infty}w_2(x_1,x_2,\tau)]\,dx_1\,dx_2. \tag{6.51}$$

If a purely random process is being examined, then when $\tau\longrightarrow\infty$ the values of ξ (t) and ξ (t $+\tau$) become independent and the two-dimensional distribution function becomes in this case equal to the product of two one-dimensional distribution functions. Then from (6.51), after separating the variables of integration, we obtain

$$a^2=\left[\int_{-\infty}^{\infty}f(x)w_1(x)\,dx\right]^2. \tag{6.52}$$

Having found the correlation function of a random process at the output of a nonlinear network, it is possible, employing (5.44), i.e., effecting a Fourier transformation, to obtain the power spectrum of this process.

If the process at the input of a nonlinear network constitutes a sum of a stationary random process and a determined process, then contrary to the case of a linear network, the passage through the network of the determined part may not be treated separately from the purely random process. The passage through the nonlinear network of both terms must be studied jointly.

The direct computation of the integral in formula (6.50) is, as a rule, very difficult. Therefore it is expedient first to transform it into a form which separates the variables of integration in the double integral. Several methods exist for such a simplification of formula (6.50).

A. Series expansion of two-dimensional probability density.

Let $w_{11}(x_1)$ and $w_{12}(x_2)$ be one-dimensional distribution functions, corresponding

to a two-dimensional probability density of $w_2(x_1, x_2, \tau)$, i.e.,

$$w_{11}(x_1)=\int_{-\infty}^{\infty} w_2(x_1, x_2, \tau)\,dx_2,$$
$$w_{12}(x_2)=\int_{-\infty}^{\infty} w_2(x_1, x_2, \tau)\,dx_1. \tag{6.53}$$

Let us take $w_{11}(x_1)$ and $w_{12}(x_2)$ as weighting functions and let us construct two aggregates of normalized orthogonal polynomials* $Q_{n1}(x_1)$ and $Q_{n2}(x_2)$, which must satisfy the conditions of orthogonality

$$\int_{-\infty}^{\infty} w_{11}(x_1)Q_{n1}(x_1)Q_{m1}(x_1)\,dx_1=\begin{cases}1 & m=n,\\ 0 & m\neq n.\end{cases} \tag{6.54}$$

$$\int_{-\infty}^{\infty} w_{12}(x_2)Q_{n2}(x_2)Q_{m2}(x_2)\,dx_2=\begin{cases}1 & m=n,\\ 0 & m\neq n.\end{cases} \tag{6.54'}$$

Then the two-dimensional distribution function $w_2(x_1, x_2, \tau)$ may be expanded into a double series in terms of these orthogonal polynomials

$$w_2(x_1, x_2, \tau)=w_{11}(x_1)\,w_{12}(x_2)\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} a_{mn}(\tau)\,Q_{m1}(x_1)\,Q_{n2}(x_2). \tag{6.55}$$

The coefficients a_{mn} can be found by multiplying both sides of equation (6.55) by $Q_{k1}(x_1)\,Q_{l2}(x_2)$ and integrating with the use of the orthogonality condition of (6.54) and (6.54'). As a result, all terms vanish, except one ($k = m$, $l = n$), and consequently

$$a_{mn}(\tau)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_2(x_1, x_2, \tau)\,Q_{m1}(x_1)\,Q_{n2}(x_2)\,dx_1dx_2. \tag{6.56}$$

In many cases of practical importance $a_{mn} = 0$ when $m \neq n$. For this class of distribution functions formulas (6.55) and (6.56) become simpler

$$w_2(x_1, x_2, \tau)=w_{11}(x_1)\,w_{12}(x_2)\sum_{n=0}^{\infty} a_n(\tau)\,Q_{n1}(x_1)\,Q_{n2}(x_2), \tag{6.57}$$

$$a_n(\tau)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_2(x_1, x_2, \tau)\,Q_{n1}(x_1)\,Q_{n2}(x_2)\,dx_1dx_2. \tag{6.58}$$

Substituting (6.57) into (6.50) and separating the variables of integration, we obtain

---

* On the construction of an aggregate of normalized orthogonal polynomials on the basis of an assigned weighting function cf. in the book of V. L. Gonchareov "Teoriya interpolirovaniya i problizheniya funktsiy" ("The Theory of Interpolation and Approximation of Functions"), Costekhizdat, 1954.

$$B(\tau)=\sum_{n=0}^{\infty} c_{1n}c_{2n}a_n(\tau), \tag{6.59}$$

where

$$c_{1n}=\int_{-\infty}^{\infty} f(x_1)Q_{n1}(x_1)w_{11}(x_1)\,dx_1, \tag{6.60}$$

$$c_{2n}=\int_{-\infty}^{\infty} f(x_2)Q_{n2}(x_2)w_{12}(x_2)\,dx_2. \tag{6.61}$$

Formula (6.59) yields the expression of the correlation function of a stationary random process at the output of a nonlinear network in the form of the series of functions $a_n(\tau)$, which are determined by the correlational characteristics of the process at the input. Such a method of computing $B(\tau)$ may be called the direct method (or the method of correlations).

If the process at the input is nonstationary, the two-dimensional probability density w_2 will be also a function of time t and therefore c_{1n}, c_{2n} and a_n will also depend on t. For determination of the correlation function, a supplemental time averaging will be necessary.

B. Contour integral method.

The second method consists in making use of the fact that the characteristics of several nonlinear networks permit representation by means of a contour integral in the form of

$$f(x)=\frac{1}{2\pi}\int_C g(iu)\,e^{ixu}\,du. \tag{6.62}$$

If (6.62) is substituted into (6.50), by changing the order of integration we find

$$B(\tau)=\frac{1}{4\pi^2}\int_{C_1} g(iu_1)\int_{C_2} g(iu_2)\times$$
$$\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_2(x_1,x_2,\tau)\,e^{i(x_1u_1+x_2u_2)}dx_1dx_2du_2du_1$$

and, employing (3.74), we obtain

$$B(\tau)=\frac{1}{4\pi^2}\int_{C_1} g(iu_1)\int_{C_2} g(iu_2)\,\Theta_2(u_1,u_2,\tau)\,du_2du_1, \tag{6.63}$$

where $\Theta_2(u_1, u_2, \tau)$ is the two-dimensional characteristic function of a random process at the input of a nonlinear network. The function $\Theta_2(u_1, u_2, \tau)$ may, analogously to (6.57) be expanded into the series

$$\Theta_2(u_1, u_2, \tau)=\Theta_{11}(u_1)\cdot\Theta_{12}(u_2)\sum_{n=0}^{\infty} M_{n1}(u_1)M_{n2}(u_2)b_n(\tau), \tag{6.64}$$

where $M_{n1}(u_1)$ and $M_{n2}(u_2)$ are polynomials of the n-th power.

Then from (6.63) we obtain the expression, analogous to (6.59)

$$B(\tau)=\sum_{n=0}^{\infty} d_{1n}d_{2n}b_n(\tau), \tag{6.65}$$

in which

$$d_{1n}=\frac{1}{2\pi}\int_{c_1} g(iu_1)M_{n1}(u_1)\Theta_{11}(u_1)\,du_1, \tag{6.66}$$

$$d_{2n}=\frac{1}{2\pi}\int_{c_2} g(iu_2)M_{n2}(u_2)\Theta_{12}(u_2)\,du_2. \tag{6.67}$$

If the process at the input is nonstationary, then the two-dimensional characteristic function Θ_2 will be also a function of time t, and therefore d_{1n}, d_{2n}, and b_n will also depend on t. For determination of the correlation function, a supplemental time averaging will be necessary.

C. The envelope method.

In problems where a narrow-band process is subjected to a nonlinear transformation, it is expedient to employ the representation of that process in the form of (6.33), taking into account (6.36)

$$\xi(t)=E(t)\cos[\omega_0 t-\varphi(t)],$$

and to represent the process at the output of the nonlinear network in the form of

$$\begin{aligned}\eta(t)=f[\xi(t)]=f_0(E)+\\ +f_1(E)\cos(\omega_0 t-\varphi)+f_2(E)\cos 2(\omega_0 t-\varphi)+\ldots+\\ +f_n(E)\cos n(\omega_0 t-\varphi)+\ldots,\end{aligned} \tag{6.68}$$

where

$$f_n(E)=\frac{\varepsilon_n}{\pi}\int_0^{\pi} f(E\cos\psi)\cos n\psi\,d\psi$$

$$(\varepsilon_0=1,\ \varepsilon_n=2,\ n\neq 0). \tag{6.68'}$$

It can be shown* that E and φ are independent, and the spectra $f_n(E)$ practically do not overlap. The correlation function of the process $\eta(t) = f\left[\xi(t)\right]$ is represented in this case by the series

$$B(\tau) = \sum_{n=0}^{\infty} B_n(\tau) \cos n\omega_0\tau, \tag{6.69}$$

in which the functions $B_n(\tau)$ are determined by the formula

$$B_n(\tau) = \int_0^{\infty}\int_0^{\infty} f_n(y_1) f_n(y_2) W_2(y_1, y_2, \tau)\, dy_1 dy_2, \tag{6.70}$$

where $W_2(y_1, y_2, \tau)$ is the two-dimensional distribution function of the envelope of the process. The envelope method was developed by V. I. Bunimovich [13].

7. Distribution Function of Random Process at Output of Nonlinear Network.

As was already noted in Section 1, for a nonlinear and noninertial network the function f(x) in (6.49) is singled-valued, i.e., the value of the random function $\eta(t)$, characterizing the process at the output of a nonlinear network at any instant in time, is determined only by the value $\xi(t)$ of the random function corresponding to the input at the same instant in time: $\eta(t) = f\left[\xi(t)\right]$. Therefore for determination of the distribution function $W_n(y_1, y_2, \ldots y_n, t_1, \ldots t_n)$ of the random process at the output of a nonlinear network, it is sufficient to have the corresponding distribution function $w_n(x_1, x_2, \ldots x_n, t_1, \ldots t_n)$ for the input, and to employ the general formula set forth in Chapter III for the replacement of the variables in distribution functions in the transformation

$$\eta_k = f(\xi_k), \quad (k = 1, 2, \ldots n), \tag{6.71}$$

where $\xi_k = \xi(t_k),\ \eta_k = \eta(t_k)$.

Sometimes these computations may be simplified, if first are found not the distribution functions themselves, but their characteristic functions.

The employment of characteristic functions for determining the distribution

---

* For proof of the independence of E and φ, it is sufficient to show that if ξ_1 and ξ_2 are independent, and normally distributed with parameters of $a_1 = a_2 = 0$, $\sigma_1 = \sigma_2 = \sigma$, then $\eta_1 = \xi_1^2 + \xi_2^2$ and $\eta_2 = \xi_1/\xi_2$ are also independent.

functions of a process at the output of a nonlinear network will become necessary, each time that function f(x) in (6.70) does not have a reciprocal (e.g., in the case of half-wave detection) and the employment of the formulas in Chapter III is impossible.

Let us employ (3.73) and write the n-dimensional characteristic function of a process at the output of a nonlinear network

$$\Theta_n(v_1, v_2, \ldots v_n, t_1, t_2, \ldots t_n) = = m_1\left\{e^{i[v_1 f(\xi_1)+v_2 f(\xi_2)+\ldots+v_n f(\xi_n)]}\right\}, \tag{6.72}$$

or

$$\Theta_n(v_1, v_2, \ldots v_n, t_1, t_2, \ldots t_n) = = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{i[v_1 f(x_1)+v_2 f(x_2)+\ldots+v_n f(x_n)]}\times \times w_n(x_1, x_2, \ldots x_n, t_1, t_2, \ldots t_n)\cdot dx_1 dx_2 \ldots dx_n. \tag{6.73}$$

Performing an inverse Fourier transformation upon Θ_n, we find the n-dimensional distribution function $W_n(y_1, y_2, \ldots y_n, t_1, t_2, \ldots t_n)$ of a process at the output of a nonlinear network.

For determining the statistical characteristics of the envelope E(t) and phase φ (t) of the narrow-band random process ξ (t), we represent it in the form of a sum (cf. p. 237)

$$\xi(t) = E(t)\cos[\omega_0 t - \varphi(t)] = A(t)\cos\omega_0 t + C(t)\sin\omega_0 t, \tag{6.74}$$

where A(t) and C(t) are independent random processes. If the n-dimensional distribution functions of these random processes are equal respectively to $w_{n1}(x_1, x_2, \ldots x_n, t_1, \ldots t_n)$ and $w_{n2}(y_1, y_2, \ldots y_n, t_1, \ldots t_n)$, then by virtue of their independence the joint distribution of these processes will be characterized by a 2n-dimensional distribution function

$$W_{2n}(x_1, \ldots x_n, y_1, \ldots y_n, t_1, \ldots t_n) = = w_{n1}(x_1, \ldots x_n, t_1, \ldots t_n)\cdot w_{n2}(y_1, \ldots y_n, t_1, \ldots t_n). \tag{6.75}$$

In order to find the distribution functions of the envelope E(t) and phase φ (t), we transform (6.75) to polar coordinates

$$x_k = r_k \cos\theta_k,$$
$$y_k = r_k \sin\theta_k. \quad (6.76)$$

After the substitution of (6.76), in place of the distribution function (6.75) of the variables $x_1, \ldots x_n$, $y_1, \ldots y_n$ there results the 2n-dimensional distribution function of the variables $r_1, \ldots r_n$, $\Theta_1, \ldots \Theta_n$, equal to

$$\begin{aligned} &w_{2n}(r_1, \ldots r_n, \theta_1, \ldots \theta_n, t_1, \ldots t_n) = \\ &= W_{2n}[r_1 \cos\theta_1, \ldots r_n \cos\theta_n, r_1 \sin\theta_1, \ldots \\ &\ldots r_n \cos\theta_n, t_1, \ldots t_n) |D_n|, \end{aligned} \quad (6.77)$$

where D_n is the jacobian of transformation (6.76), equal to

$$D_n = \frac{\partial(x_1, \ldots x_n, y_1, \ldots y_n)}{\partial(r_1, \ldots r_n, \theta_1, \ldots \theta_n)} = r_1 \cdot r_2 \ldots r_n. \quad (6.78)$$

The distribution function of the envelope is obtained by n-fold integration of (6.77) with respect to variables $\Theta_1, \ldots \Theta_n$

$$\begin{aligned} &w_n(r_1, r_2, \ldots r_n, t_1, \ldots t_n) = \\ &= r_1 \cdot r_2 \ldots r_n \int_0^{2\pi}\int_0^{2\pi} \ldots \int_0^{2\pi} w_{n1}(r_1 \cos\theta_1, r_2 \cos\theta_2, \\ &\ldots r_n \cos\theta_n, t_1, \ldots t_n) \times \\ &\times w_{n2}(r_1 \sin\theta_1, r_2 \sin\theta_2, \ldots r_n \sin\theta_n, t_1, \ldots t_n) d\theta_1 \times \\ &\times d\theta_2, \ldots d\theta_n. \\ &r_i > 0 \ (i = 1, 2, \ldots n) \end{aligned} \quad (6.79)$$

The distribution function of phase $\varphi(t)$ of the random process is obtained from (6.77) by integration with respect to variables $r_1, r_2, \ldots r_n$

$$\begin{aligned} &w_n(\theta_1, \theta_2, \ldots \theta_n, t_1 \ldots t_n) = \\ &= \int_0^{\infty}\int_0^{\infty} \ldots \int_0^{\infty} r_1, r_2 \ldots r_n w_{n1}(r_1 \cos\theta_1, r_2 \cos\theta_2, \ldots \\ &\ldots r_n \cos\theta_n, t_1, \ldots t_n) w_{n2}(r_1 \sin\theta_1, r_2 \sin\theta_2, \ldots \\ &\ldots r_n \sin\theta_n, t_1, \ldots t_n) dr_1 dr_2 \ldots dr_n, \\ &-\pi < \theta_i < \pi \ (i = 1, 2, \ldots n). \end{aligned} \quad (6.80)$$

In certain problems a result may be arrived at more rapidly with the aid of the characteristic function. Since in (6.74) the random processes A(t) and C(t) are

independent, the characteristic function $\Theta_n(\nu_1, \nu_2, \ldots \nu_n, t_1, \ldots t_n)$, of the process $\eta(t)$ obtained as a result of the nonlinear transformation (6.49), may in accordance with (3.73) be computed according to the formula

$$\begin{gathered}\Theta_n(v_1, \ldots v_n, t_1, \ldots t_n) = \\ = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{2n \text{ times}} e^{iv_1 f\left(\sqrt{x_1^2+y_1^2}\right) + \ldots + iv_n f\left(\sqrt{x_n^2+y_n^2}\right)} \times \\ \times w_{n1}(x_1, \ldots x_n, t_1, \ldots t_n) \times \\ \times w_{n2}(y_1, \ldots y_n, t_1, \ldots t_n) dx_1 \ldots dx_n dy_1 \ldots dy_n\end{gathered} \tag{6.81}$$

8. On the Determination of the Distribution Function of a Random Process at the Output of a Linear Network

As we have seen in §2, the problem of the transformation of the correlation function of a process and its spectrum is sufficiently simple in a linear network and does not even require a knowledge of the distribution functions of the input process. An incomparably more complicated problem is the determination of the distribution function of process $\xi_2(t)$ at the output of the linear network.

Only in one special case, when the process $\xi_1(t)$ at the input of the linear network is normal, can this problem be solved simply. The random process $\xi_2(t)$ is the limit of the integral sum $\sum_{k=0}^{N} \xi_1(t-\tau_k) h(\tau_k)(\tau'_{k+1}-\tau'_k)$, where $\tau'_k < \, < \tau_k < \tau'_{k+1}$, when $|\tau'_{k+1}-\tau'_k| \to 0$ [cf. (6.3)]. The random variables $\xi_1(t)$, $\xi_1(t-\tau_1), \ldots \xi_1(t-\tau_n)$ in the case at hand are linked by an N + 1 -dimensional normal law of distribution. Since a sum of normally distributed random variables (even of dependent ones, cf. Section 9, Ch. 3) is normally distributed, then, consequently, the sum in question has a normal distribution for any N.

Thus, in passage through linear networks, the normal random process retains its distribution functions, i.e., remains normal. Changes occur only in the correlation function and power spectrum, in accordance with the formulas indicated above.

It has already been noted in Section 1, that an analysis typical of radio apparatus is one of a linear network following a nonlinear element (Fig. 43). Therefore even in

those cases when a normal random process is subjected to a nonlinear transformation, this process at the input of the succeeding linear network is no longer normal.

The problem of the transformation of a distribution function in a linear network, at the input of which there acts a random process differing from the normal, is an extremely difficult one. There exist several approximate methods of solving this problem, each of which is based on special assumptions with respect to the statistical characteristics of the input random process and the properties of the linear network itself[15].

In Chapter IX, this problem will be examined under the assumption that the process at the input of the linear network is the square of a normal random process. Here we shall pause merely for one general approximate method of determining one-dimensional statistical characteristics, which consists in the computation of the distribution moments m_k of a process at the output of a linear network [9],[10]. Since th one-dimensional characteristic function $\Theta(v)$ may be represented by a series [cf. (3.67)]

$$\Theta(v)=\sum_{k=0}^{\infty}\frac{i^k m_k v^k}{k!},$$

therefore, having a sufficient number of distribution moments, it is possible with a certain degree of approximation also to synthesize a distribution function or, at least, to have a conception of its type.

If there were known the n-dimensional correlation function of the random process $\xi_2(t)$ at the output of a linear network, i.e., the mixed distribution moment of the n + 1-th order

$$B_n(\tau_1,\ldots\tau_n,t)=m_1\{\xi_2(t)\xi_2(t+\tau_1)\ldots\xi_2(t+\tau_n)\}=$$
$$=\underbrace{\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}}_{n+1\ \text{times}}x_1\ldots x_{n+1}w_{n+1}(x_1,\ldots x_{n+1},\tau_1,\ldots$$
$$\ldots\tau_n,t)\,dx_1\ldots dx_{n+1},$$

then the n + 1-th initial moment of the one-dimensional law of distribution of this process would be determined by means of the limiting relationship

$$m_{n+1}(t)=\int\limits_{-\infty}^{\infty} x^{n+1} w_1(x, t)\,dx=$$

$$=m_1\{\xi_2^{n+1}(t)\}=\lim_{\substack{\tau_1\to 0\\ \cdots\\ \tau_n\to 0}} B_n(\tau_1,\ldots\tau_n, t). \qquad (6.82)$$

Let us designate by $F_n(\omega_1,\ldots\omega_n)$, the spectrum (Fourier transformation) of the n-dimensional correlation function $B_n^*(\tau_1,\ldots\tau_n)$, averaged over time. Then

$$B_n^*(\tau_1,\ldots\tau_n)=$$

$$=\frac{1}{(4\pi)^n}\underbrace{\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}}_{n\ \text{times}} F_n(\omega_1,\ldots\omega_n)\,e^{i(\omega_1\tau_1+\ldots+\omega_n\tau_n)}\,d\omega_1\ldots d\omega_n. \qquad (6.83)$$

From (6.82) and (6.83) there follows

$$m_{n+1}=\lim_{\substack{\tau_1\to 0\\ \cdots\\ \tau_n\to 0}} B_n^*(\tau_1\ldots\tau_n)=$$

$$=\frac{1}{(4\pi)^n}\underbrace{\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}}_{n\ \text{times}} F_n(\omega_1,\ldots\omega_n)\,d\omega_1\ldots d\omega_n. \qquad (6.84)$$

Thus, the n + 1-th initial moment of the one-dimensional distribution function of a stationary process (or the mean value of this moment for a nonstationary process) is equal to the full volume of the spectrum of the time-averaged, n-dimensional correlation function of this process.

Consequently, in order to determine the n + 1-th initial moment of the one-dimensional distribution function of a process at the output of a linear network it is sufficient to know for this process the indicated n-dimensional spectrum $F_n(\omega_1, \ldots\omega_n)$. It may be found by means of the transformation formula for such spectra in linear networks:

$$F_n^{(2)}(\omega_1,\ldots\omega_n)=F_n^{(1)}(\omega_1,\ldots\omega_n)\cdot k\left(-i\sum_{m=1}^{n}\omega_m\right)\prod_{m=1}^{n}k(i\omega_m), \qquad (6.85)$$

which constitutes a generalization of formula (6.4) for the one-dimensional case.

Here $k(i\omega)$ is the transfer function of the linear system.

It should be noted, that in those cases when the spectrum band of the process at the input of a linear network is much wider than the pass band of this network, the process at the output always has a tendency toward normalization, i.e., its distribution functions are sufficiently closely approximated by normal ones. This situation is a direct consequence of the central limit theorem of Lyapunov, set forth in Section 1 of Ch. IV. In fact, let us replace the process at the input of a linear network by a sequence of base pulses, the duration ΔT of each of which is much less than $\frac{1}{\Delta f}$, where Δf is the pass bandwidth of the network. The process at the output of the linear network after the passage of such a sequence of base pulses will in practice differ little from the process which takes place, when at the input there acts the original random process. The reaction of the linear network to a sequence of base pulses may be represented in the form of a sum

$$\eta(t)=\sum_k h(t-k\Delta T)\xi(k\Delta T)\,\Delta T. \tag{6.86}$$

If the power spectrum of the initial random process $\xi(t)$ is so wide, that the correlation time τ_0 of this process is much smaller than ΔT, then any two base-pulse amplitudes $\xi(k\Delta T)$ and $\xi(r\Delta T)$ $(k \neq r)$ will be independent. It can then be seen from (6.86), that the process at the output of a linear network is the sum of a large number of independent random variables, which in accordance with the Lyapunov theorem should have a distribution close to the normal.

Since the product of the effective width of the spectrum of a random process ΔF by the correlation time τ_0 is on the order of unity, therefore from the conditions cited above of the normalization of a process at the output of a linear network

$$\tau_0 \ll \Delta T \ll \frac{1}{\Delta f}$$

it follows, that

$$\frac{1}{\Delta F} \ll \frac{1}{\Delta f}, \text{ or } \frac{\Delta f}{\Delta F} \ll 1, \tag{6.87}$$

i.e., for the normalization of a process it is necessary, that the band of a linear network be much narrower than the effective width of the power spectrum of the random process acting on the input.

Of course, the considerations cited here do not constitute a rigorous proof of the theorem of the normalization of a process, when $\frac{\Delta f}{\Delta F} \longrightarrow 0$. For such a proof it would be necessary to take into account the presence of correlation links between the terms of sum (6.86), and to expand the central limit theorem to the sum of dependent random variables (cf. S. N. Bernshteyn. Teoriya veroyatnostey (Probability theory), 4-th edition. Gostekhizdat, 1946.)

Literature

1. A. M. Yaglom. Vvedeniye v teoriyu statsionarnykh sluchaynykh funktsiy (Introduction to the theory of stationary random functions), "Uspekhi matematicheskikh nauk", 1952, V. VII, No. 5.

2. V. V. Solodovnikov. Vvedeniye v statisticheskuyu dinamiku sistem avtomaticheskogo upravleniya (Introduction to the statistical dynamics of automatic control systems). Gostekhizdat, 1952.

3. N. Wiener. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. John Wiley, New York 1949.

4. G. A. Levin and B. P. Levin. O polose propuskaniya lineynoy sistemy i neickazhennom vosproizvedenii signalov (On the pass band of a linear network and the undistorted reproduction of signals). "Radiotekhnika", 1954, V. 9, No. 2.

5. J. F. Barrett and D. G. Lampard. An Expansion for some Second-order Probability Distributions and its Application to noise Problems. IRE Transaction (sic.), IT-1, No. 1, March 1955.

6. D. A. Bell. Wiener's Theory of Linear Filtering. Wireless Engineer, June 1953.

7. K. S. Berkovitz. Optimum Linear Shaping and Filtering Networks. Jours. Appl. Phys., 24, No. 8, 1953.

8. Ya. S. Itskhoki. Nelineynaya radiotechnika (Nonlinear radio engineering), Moskva,

"Sovetskoye Radio", 1955.

9. B. R. Levon and S. A. Smirnov. K voprosu o normalizatsii sluchaynogo protsessa pri prokhozhdenii cherez uzkopolosnuyu sistemu (On the question of the normalization of a random process in passage through a narrow-band network). Vestnik NII, No. 2 (47), 1954.

10. P. I. Kuznetsov, R. L. Stratonovich, V. I. Tikhonov. Prokhozhdeniye nekotorykh sluchaynykh funktsiy cherez lineynyye sistemy (The passage of some random functions through linear networks). "Avtomatika i telemekhanika" (Automation and telemechanics), 1953, V. XIV, No. 4.

11. P. I. Kuznetsov, R. L. Stratonovich, V. I. Tikhonov. Prokhozhdeniye sluchaynykh funktsiy cherez nelineynyye sistemy (The passage of random function through nonlinear networks.) "Avtoma ika i Telemekhanika", 1953, V. XIV, No. 4.

12. B. E. Keise. The linear, Input-controlled, Variable-pass Network. IRE Trans. Inform. Theory, IT-1, no. 1, 1955.

13. V. I. Bunimovich. Fluktuatsionnyye protsessy v radiopriyemnykh ustroystvakh (Fluctuation processes in radio receivers). Moskva, "Sovetskoye Radio", 1951.

14. S. O. Rice. Mathematical Analysis of Random Noise, BSTJ, 23, No. 3, July 1944, 24, No. 1, Jan. 1945. (Translation [into Russian] available in the collection "Teoriya peredachi èlektricheskikh signalov pri nalichii pomekh" (Theory of the transmission of electrical signals in the presence of noise), Moskva, For. Lit. Pub. Hse., 1953).

15. Rosenbloom, Heilfron, Trautman. Analysis of linear systems with randomly varying input parameters. IRE Convent. Rec., 3, No. 4, 1955.

16. Siegert A. I. F. Passage of Stationary Processes through Linear and Nonlinear Devices. Trans. IRE, PGIT-3, 1954.

17. L. A. Zadeh, J. R. Ragazzini. Optimum filters for detection of signal in noise, PIRE, 40, 1952.

18. Ya. I. Samoylenko. Perekhodnyye protsessy v lineynykh tsepyakh pri vklyuchenii neregulyarnykh e.d.s. (Transient processes in linear networks when irregular e.m.f's

Radiotekhnika, 1956, No. 6.

19. Tszyan' Syue - Sen'. Technical cybernetics. Moskva. For. Lit. Pub. Hse. 1956.